VNIVERSITAT DE VALÈNCIA

Departament d'Informàtica

No Linealidades en la Percepción del Contraste y Aplicaciones en Restauración de Imágenes y Estimación del Movimiento.

Tesis Doctoral

Juan Gutiérrez Aguado

Valencia, Julio de 2005

JESÚS MALO LÓPEZ Profesor Titular de Universidad del Departament d'Òptica de la Universitat de València y FRANCESC JOSEP FERRI RABASA, Profesor Titular de Universidad del Departamet d'Informàtica de la Universitat de València

CERTIFICAN:

Que la presente memoria, *No Linealidades en la Percepción del Contraste y Aplicaciones en Restauración de Imágenes y Estimación del Movimiento*, resume el trabajo de investigación realizado bajo su dirección en la Escola Tècnica Superior d'Enginyeria de la Universitat de València por Juan Gutiérrez Aguado y constituye su Tesis por Compendio de Publicaciones para optar al grado de Doctor en Ingeniería Informática.

Y para que así conste presentan la referida memoria, firmando el presente certificado en Valencia a 20 de Julio de 2005

Fdo: Jesús Malo López

Fdo. Francesc J. Ferri Rabasa

No Linealidades en la Percepción del Contraste y Aplicaciones en Restauración de Imágenes y Estimación del Movimiento.

Juan Gutiérrez Aguado

Memoria presentada para optar al grado de
Doctor en Ingeniería Informática

Trabajo realizado en el Departament d'Informàtica
Escola Tècnica Superior d'Enginyeria
Universitat de València

Directores

Jesús Malo López

Francesc J. Ferri Rabasa

Agradecimientos

He recorrido un largo camino desde que comencé mi actividad investigadora. Durante este tiempo he sido bastante afortunado, he tenido la suerte de trabajar en un ambiente estimulante, con compañeros de trabajo sin los cuales habría sido difícil llegar hasta aquí.

Quiero agradecer a Francesc Ferri el haberme brindado la oportunidad de trabajar en este campo. Quiero agradecer su tiempo, confianza y que me ofreciera la posibilidad de ir al Digital Imaging Research Center de la Kingston University (UK). Quiero agradecer a Graeme Jones y especialmente a Paolo Remagnino su hospitalidad durante la citada estancia.

La doctora Irene Epifanio (ahora en la Universitat Jaume I) ha sido parte esencial en algunos de los artículos que se presentan en esta tesis.

La idea de introducir un operador perceptual para la regularización de la restauración de imágenes degradadas comenzó a tomar forma durante una estancia con el Prof. Per Christian Hansen en la Technical University of Denmark. Es de agradecer la actitud del Prof. Hansen admitiendo a un aventurado que le envió un correo indicando que quería trabajar en la regularización del flujo óptico. Ha cambiado la aplicación, pero las discusiones sobre la L-curva me han sido de gran utilidad.

A Jesús Malo quiero agradecerle su confianza, su tiempo, su entrega, su paciencia y su humor. Sin él éste trabajo simplemente no existiría

Quiero manifestar que el Departament d'Informàtica ha sido un espacio de libertad donde crecer, madurar y pasar algo más de ocho horas al día, y por ello quiero darle las gracias a todos sus miembros.

A mi familia que siempre me ha apoyado en esta aventura.

Finalmente quiero agradecer a María José su valentía, su comprensión, su apoyo y el haberme permitido disponer de un tiempo que sin duda le correspondía.

Por supuesto no me olvido de Carlos que en el momento de escribir esta memoria tiene ya 9 meses y es lo único que me quita el sueño...

ÍNDICE GENERAL

1. INTRODUCCIÓN

Las imágenes de escenas naturales (imágenes fotográficas no sintéticas) presentan un comportamiento estadístico muy particular. Específicamente, representando cada imagen (acromática) de $N \times M$ píxels como un punto en un espacio de $N \times M$ dimensiones, resulta que la densidad de probabilidad de las escenas naturales en dicho espacio de imágenes es extremadamente no uniforme.

El conocimiento de la naturaleza estadística de estos estímulos puede ser de gran utilidad en sistemas de procesado de imágenes y visión artificial. Por citar algunos ejemplos, cuando se considera una imagen fotográfica que ha sido degradada por algún tipo de ruido, el proceso de restauración puede entenderse como la búsqueda de la imagen con características estadísticas naturales que esté más próxima a la observación ruidosa. El problema de la codificación de imágenes esencialmente se reduce a utilizar una mayor proporción de los bits disponibles para codificar aquellas regiones del espacio de imágenes que estén más densamente pobladas. Y problemas tales como el incremento de resolución o la síntesis de imágenes involucran la selección de una imagen (una muestra) de alta probabilidad de la distribución de imágenes naturales introduciendo quizá algún conjunto de restricciones. Por este motivo el desarrollo de modelos que revelen características estadísticas típicas de las imágenes es de gran utilidad en ese tipo de aplicaciones.

No obstante, a la hora de abordar el problema de la adecuada representación de las señales para la solución de problemas de procesado de imágenes (que es equivalente a la descripción de las características estadísticas de las imágenes), es posible otra aproximación diferente.

Los sistemas biológicos de visión están expuestos de forma continua a este tipo de señales y gracias al desarrollo evolutivo han organizado sus sensores para resolver de forma muy eficiente problemas de procesado de imágenes. La comprensión del funcionamiento, las reglas de organización de dichos sistemas y la relación de estos con la estadística de las señales que deben procesar puede resultar muy útil en aplicaciones como las citadas arriba.

En esta Tesis adoptamos esta segunda aproximación en una doble vertiente: (1) profundizando en la interpretación del comportamiento de los sensores corticales del sistema visual humano según la estadística de las imágenes naturales,

y (2) utilizando los modelos paramétricos de este comportamiento como alternativas a la modelización estadística de la señal en aplicaciones de restauración de imágenes, codificación y estimación eficiente del movimiento.

En primer lugar, consideremos esquemáticamente las etapas tempranas del proceso de visión en los mamíferos para centrar cuales son los problemas abiertos en cuanto a la comprensión de su funcionamiento y la relación de este funcionamiento con ciertos problemas de procesado de imágenes.

Este proceso comienza en las neuronas situadas en la retina que transforman la luz incidente en estímulos eléctricos que son enviados a través de los núcleos laterales geniculados (NLG) hacia el cortex visual, donde es analizada en primera instancia por el área V1. En este camino, el estímulo sigue diferentes rutas siendo transformado y procesado en paralelo. La información extraída de cada una de estas rutas es analizada en etapas posteriores de forma independiente [1], es decir, cada una de estas rutas contiene información que puede ser utilizada para tomar decisiones independientemente del resto de las rutas.

Las propiedades de los sensores de V1 han sido estudiadas siguiendo los trabajos de Hubel y Weisel [2] en los que determinaron cuales eran los estímulos que excitan óptimamente a las neuronas de esta región. Encontraron dos tipos de neuronas a las que denominaron simples y complejas. Las primeras responden ante barras de una determinada orientación, escala y colocadas en una posición específica en el campo visual y las segundas responden ante barras de una determinada orientación pero pueden aparecer en cualquier posición del campo visual.[1]

Otra característica notable que presentan las neuronas de V1 es que responden no linealmente al contraste del estímulo [4]. Además de esta saturación en la respuesta, la respuesta de una neurona al estímulo óptimo se ve afectada por la presencia de otros estímulos [5]: la respuesta disminuye en presencia de otros estímulos que se conocen como estímulos de enmascaramiento.

Una vez conocidas las propiedades de las etapas iniciales de los sistemas sensoriales es posible proponer modelos empíricos que sistematicen dichas propiedades. El modelo más simple es un modelo lineal en el que se propone un análisis de la imagen utilizando un banco de filtros pasa banda [6, 7, 8, 9] de tipo *wavelet* [10] y la aplicación de una ganancia dependiente de la frecuencia a cada uno de los coeficientes obtenidos [11]. Los modelos más recientes incorporan una no linealidad tras el banco de filtros que puede ser puntual [12, 13] o que puede involucrar a los coeficientes vecinos [4, 14, 15, 5].

Esta enumeración de las propiedades de las neuronas no responde sin em-

[1] Como en tantas otras ocasiones estos descubrimientos fueron casuales ya que los estímulos diseñados para excitar a la neurona objetivo en el cortex visual del gato no producían ninguna excitación. En uno de los experimentos, la neurona respondió al borde del cristal (en el que se encontraba el estímulo) cuando era introducido en su soporte [3].

bargo a las preguntas: ¿por qué los sistemas sensoriales tienen una determinada estructura?, de todas las posibles estructuras que pueden surgir a partir de los elementos disponibles sólo unas cuantas están presentes, ¿son las estructuras generadas óptimas o al menos favorecidas sobre otras?, ¿respecto a qué objetivo?. Si somos capaces de encontrar este principio, estaríamos describiendo estas primeras etapas como el modo en el que la biología consigue lograr dicho objetivo.

La respuesta a esta pregunta se puede abordar desde dos perspectivas [16]. La primera consiste en la investigación de cómo mecanismos de bajo nivel (a nivel de neuronas) pueden general los tipos de estructuras neuronales que se encuentran en las primeras etapas de los sistemas perceptuales. Esta aproximación conlleva modelar la formación y la modificación de las conexiones neuronales mediante reglas simples (por ejemplo reglas tipo Hebb para ajustar los pesos entre conexiones), expresar estas reglas matemáticamente y realizar simulaciones para comprobar qué tipo de estructuras generan. El tipo de estructuras que se forman y que han sido estudiadas incluyen campos receptivos selectivos a la orientación [17]. La segunda aproximación consiste en una serie de ideas sobre cómo se pueden codificar y transformar los datos en un sistema perceptual y para qué pueden servir estas transformaciones desde el punto de vista de la teoría de la información, es lo que se conoce como la *hipótesis de la codificación eficiente* [18, 19, 20, 21] (también conocida como hipótesis de Barlow).

La hipótesis de la codificación eficiente proporciona una conexión entre las propiedades estadísticas de las imágenes naturales y la estructura de los sistemas visuales biológicos expuestos a estos estímulos. En esta hipótesis se propone un principio simple de autoorganización para las neuronas sensoriales: deben organizar sus conexiones de forma que la información sensorial de entrada se codifique para maximizar la independencia estadística de sus salidas (y por tanto se minimize la redundancia).

Hay dos metodologías básicas para probar y refinar la hipótesis de la codificación eficiente, (1) una aproximación directa que consiste en examinar las propiedades estadísticas de las respuestas sensoriales frente a un conjunto de estímulos naturales, o (2) utilizar propiedades estadísticas de las imágenes naturales para derivar modelos de procesado sensorial temprano.

Esta segunda metodología ha dado lugar a un nuevo campo de investigación cuyo objetivo es determinar si las propiedades de los sistemas sensoriales son explicadas por la estadística de los estímulos a los que están expuestos. Esta aproximación presupone que existe una serie de propiedades estadísticas compartidas por la mayoría de las escenas y que mediante la caracterización de estas regularidades es posible obtener conocimiento sobre el funcionamiento del sistema visual.

Siguiendo esta aproximación se han deducido exitosamente la propiedades li-

neales del sistema visual [22, 20, 23], si embargo respecto a las propiedades no lineales se han obtenido resultados equivalentes pero imponiendo la forma funcional de los modelos empíricos [24, 25, 26]. Este conjunto de resultados sugiere que el sistema visual está adaptado a la estadística de los estímulos. Por tanto parece razonable pensar que los modelos de percepción podrían sustituir a los modelos estadísticos de los estímulos en aplicaciones de procesado de imágenes (por ejemplo para diseñar cuantizadores [27] o como modelo de imagen en restauración [28]).

1.1. Aportaciones presentadas en esta Tesis

En el contexto descrito en la sección anterior, las aportaciones presentadas en esta Tesis son tres:

1. **La saturación de la respuesta de V1 y las interacciones de enmascaramiento emergen directamente de los estímulos naturales (Trabajo A)**. Se ha mostrado que las propiedades no-lineales mas relevantes de los sensores de V1 emergen directamente de los estímulos naturales aplicando un Análisis en Componentes Independientes (ICA) no-lineal original. Desde un punto de vista teórico esto supone un resultado adicional (y con menos asunciones que otros trabajos previos) en favor de la hipótesis de la codificación eficiente, y desde un punto de vista técnico, el ICA no-lineal propuesto puede servir de herramienta para el análisis global de señales no-gaussianas.

2. **La utilización de la entropía perceptual en la estimación del movimiento genera un flujo óptico más robusto (Trabajos B, C, y D)**. Se ha profundizado en el análisis de los efectos de la inclusión de modelos de respuesta perceptual no-lineal en el diseño de cuantizadores para codificación de vídeo. En concreto, por una parte, se ha mostrado que el uso de un criterio de control de la estimación adaptativa de movimiento basado en la entropía perceptual (que hace uso del modelo de respuesta no-lineal), proporciona un flujo óptico más robusto y más adecuado para la segmentación de objetos basada en movimiento. Por otra parte, se ha analizado la importancia relativa de los avances en la estimación de movimiento y el diseño de cuantizadores en la calidad de las secuencias codificadas por sistemas de compresión de video basados en la compensación del movimiento y la cuantizacion de la transformada.

3. **La respuesta no lineal puede utilizarse como alternativa a los modelos estadísticos de segundo orden en restauración basada en regularización (Trabajo E)**. Se han utilizado modelos de respuesta per-

ceptual no-lineal para definir operadores de penalización para la restauración de imágenes degradadas en el contexto de la teoría de la regularización. La aproximación (perceptual) utilizada para el diseño de los operadores es alternativa a la modelización estadística de las imágenes en el sentido de que el modelo perceptual utilizado puede recoger propiedades que van más allá de las que pueden capturar los modelos estadísticos de segundo orden (basados en la autocorrelación).

De esta forma, se ha continuado con trabajo desarrollado previamente dentro del grupo de investigación pero también se han abierto nuevas líneas que pueden extenderse en un futuro próximo.

1.2. Organización de la memoria

En el capítulo 2 se realiza una revisión de los diferentes modelos empíricos que recogen las propiedades de las primeras etapas del procesado visual. En el capítulo 3 se presenta el trabajo realizado sobre la obtención de propiedades de las no linealidades de la neuronas de V1 utilizando la hipótesis de la codificación eficiente. El capítulo 4 muestra el trabajo realizado para la evaluación de las características del flujo óptico en el contexto de la codificación de vídeo al introducir aspectos perceptuales. En el capítulo 5 se describe el trabajo realizado en el campo de la restauración de imágenes en el marco de la regularización en el que se propone la utilización de operadores de penalización basados en un modelo perceptual. Finalmente, el capítulo 6 está dedicado a las conclusiones. Como apéndice se adjuntan los cinco trabajos que forman el compendio de publicaciones de esta Tesis.

1.3. Descripción de los Trabajos

El orden de exposición de los trabajos que conforman esta Tesis viene determinado por la agrupación en tres bloques: estudio relacionado con la hipótesis de la codificación eficiente (Trabajo A), estudio relacionado con la estimación del movimiento en el contexto de la codificación de video (Trabajos C, B y D) y estudio relacionado con la restauración de imágenes (Trabajo E).

Trabajo A **V1 non-linear properties emerge from local linear ICA**. Jesús Malo, **Juan Gutiérrez** y José Rovira. En proceso de revisión en *Network: Computation in Neural Systems*

Este trabajo de naturaleza teórica muestra cómo a partir de la hipótesis de la codificación eficiente [18, 19] se pueden reproducir propiedades de las no

linealidades observadas en neuronas de la región V1. En este sentido este trabajo extiende, por resaltar dos trabajos representativos que se basan en la hipótesis de Barlow, los resultados de B. Olshausen y D. Field [20] donde muestran que un algoritmo que aprende códigos dispersos (*sparse*) a partir de un conjunto de imágenes naturales produce como resultado funciones base lineales similares a los campos receptivos de V1 y por otra parte los resultados de O. Schwartz y E. Simoncelli [24] donde muestran que partiendo de una no linealidad concreta (el modelo de normalización divisiva) y ajustando los parámetros de esta no linealidad de forma que las respuestas resultantes sean independientes se reproducen propiedades de las no linealidades observadas en V1.

En nuestro trabajo, a diferencia de [24], no asumimos ninguna forma funcional para la no linealidad, sino que las curvas de respuesta emergen directamente de los datos (imágenes o estímulos).

Además, desde el punto de vista técnico, se propone una nueva forma de resolver el ICA no lineal. En este sentido el trabajo extiende los trabajos de Karhunen [29] y Leen [30] en los que para analizar densidades de probabilidad (PDFs) no Gaussianas utilizan representaciones lineales locales (ICAs y PCAs respectivamente) basados en una preclasificación de los datos en agrupaciones sin llegar a obtener una única representación global. En nuestra aproximación utilizamos un método diferencial para conectar las diferentes representaciones ICA locales en una única representación no lineal global.

Trabajo B **Importance of quantizer design compared to optimal multigrid motion estimation in video coding.** Jesús Malo, Francesc Ferri, **Juan Gutiérrez** e Irene Epifanio. *IEE Electronics Letters*, Vol. 36, 9, pp. 807-809, 2000

En este trabajo se evalúa la importancia relativa entre el cuantizador y la estimación del movimiento en el contexto de la codificación de video. La estimación del flujo óptico y la cuantización de la transformada DCT del error de predicción juegan papeles complementarios en esquemas de codificación de video con compensación de movimiento. En este trabajo se muestra que en cuanto al efecto visual de la secuencia reconstruida son más importantes las mejoras en el cuantizador que una estimación óptima del flujo óptico.

Trabajo C **Perceptually weighted optical flow for motion-based segmentation in MPEG-4.** Jesus Malo, **Juan Gutiérrez**, Irene Epifanio y Francesc Ferri. *IEE Electronics Letters*, Vol 36, No. 20, pp. 1693-1694, 2000

En este trabajo se muestra que el uso de la entropía perceptual [31, 32] (basado en un modelo de respuesta no lineal puntual) en algoritmos jerárquicos

de estimación del flujo óptico, produce soluciones más robustas y con más significado: las segmentaciones a partir del flujo están más correlacionadas con los objetos en movimiento en las escenas. La medida de entropía perceptual involucra una cuantización alternativa a los cuantizadores obtenidos estadísticamente [33, 34].

Para ello, se evalúan diversos métodos de estimación del flujo óptico en el contexto de la codificación de video atendiendo a su utilidad para la segmentación. Estos métodos son: flujo basado en la correspondencia de bloques uniformes [35], flujo adaptativo basado en la correspondencia de bloques basado en un criterio de división que utiliza la entropía de la imagen de error en el dominio espacial (criterio definido por F. Dufaux y F. Moscheni [36]) y flujo adaptativo basado en la correspondencia de bloques basado en un criterio de división que utiliza la entropía perceptual de la señal de error (criterio definido por J. Malo *et al.* [37]).

Los experimentos involucran métodos estándar de cálculo de *Video Object Planes* (VOPs) [38, 39], y la proposición de una función de mérito original (basada en la distancia *chamfer* [40]) para evaluar la calidad de las segmentaciones.

Trabajo D **Perceptual feed-back in multigrid motion estimation using an improved DCT quantization.** Jesús Malo, **Juan Gutiérrez**, Irene Epifanio, Francesc Ferri y José María Artigas *IEEE Transactions on Image Processing* Vol.10, 10, pp. 1411-1427, 2001

Este trabajo presenta un esquema general de codificación de video con compensación del movimiento en el que se introducen aspectos perceptuales tanto en la estimación del movimiento como en el cuantizador. Se ofrece una visión de conjunto de diversos resultados parciales que habían sido publicados con anterioridad (flujo óptico [37] y Trabajos C y B, y codificación de imágenes estáticas [27, 41]) junto con nuevos experimentos no presentados anteriormente.

Trabajo E **Regularization operators for natural images based on nonlinear perception models. Juan Gutiérrez**, Francesc Ferri y Jesús Malo *IEEE Transactions on Image Processing* (aceptado en noviembre de 2004)

En este trabajo se define una clase de operadores de penalización basados en el modelo de respuesta no lineal con interacción entre coeficientes en el contexto de la restauración de imágenes degradadas (mediante emborronamiento y adición de ruido) cuando se utiliza como método de resolución la regularización de Tikhonov [42] junto con el método de la L-curva [43].

Estos operadores se han comparado con otros ya clásicos y no adaptativos (como son el operador basado en la segunda derivada y un operador basado en la Función de Sensibilidad al Contraste) y con operadores adaptativos

basados en modelos autorregresivos (de diferente orden) siempre dentro de la formulación indicada anteriormente.

Los operadores son analizados desde diversos puntos de vista:

1. en la tarea de restauración con diferentes imágenes y con diferentes niveles de degradación (desde una degradación que consiste casi exclusivamente en emborronamiento hasta una degradación que consiste casi exclusivamente en ruido), además se muestra el comportamiento en imágenes naturalmente degradadas (tomadas con una cámara digital en diferentes condiciones de iluminación).
2. se introduce la noción de sensibilidad del operador frente al parámetro de regularización y se evalúa la sensibilidad de los diferentes operadores adaptativos,
3. se introduce la noción de robustez de un operador frente a la degradación y se evalúa la robustez de los diferentes operadores.

Los experimentos realizados muestran que los operadores propuestos funcionan bajo un amplio rango de degradaciones y pueden resultar una alternativa válida frente a operadores basados en la estadística de segundo orden.

2. MODELOS DE RESPUESTA AL CONTRASTE Y MÉTRICAS INDUCIDAS

En esta sección se revisará el modelo general de respuesta al contraste ante un estímulo. Este modelo implica una secuencia de transformaciones a diferentes dominios. Se revisarán tres modelos de respuesta: un modelo lineal, un modelo no lineal sin interacción entre coeficientes y un modelo no lineal con interacción entre coeficientes. Se mostrará cómo el modelo de respuesta induce una métrica en cada uno de los dominios.

El modelo general se utiliza en el Trabajo A y diferentes modelos funcionales de no linealidad se utilizarán en el resto de las publicaciones.

2.1. Modelo general de respuesta

Los modelos de respuesta de visión temprana frente al contraste comprenden dos etapas [5, 14, 15]. En la primera, la señal es analizada en un dominio de frecuencia local mediante un conjunto de filtros pasa banda. En los modelos más antiguos, la segunda etapa es simplemente una ganancia aplicada a cada coeficiente. En cambio, los modelos más recientes aplican una no linealidad que puede involucrar a un único coeficiente o a una serie de coeficientes, según una forma funcional conocida genéricamente como normalización divisiva [4]. Podemos describir estas dos etapas como[1]

$$\mathbf{i} \xrightarrow{\mathbf{W}} \mathbf{c} \xrightarrow{R} \mathbf{r} \tag{2.1}$$

donde el vector $\mathbf{i}$ representa a la imagen, $\mathbf{W}$ es un banco de filtros pasa banda con respuestas impulsionales localizadas espacio frecuencialmente. Cada fila de esta matriz representa la función campo receptivo de cada sensor (el tipo de transformadas utilizadas habitualmente en la literatura incluyen transformadas

[1] Debido a que en cada trabajo se utiliza una notación diferente, a la hora de redactar la memoria se presentaban dos alternativas: utilizar en cada capítulo de la memoria una notación diferente que se correspondiera con lo presentado en los trabajos, o por el contrario, intentar unificar la notación con el fin de facilitar la lectura de la memoria a costa de que la notación no se correspondiera con los trabajos. La opción elegida ha sido la segunda.

wavelet [6, 7, 8, 9] o ICA [44] y aproximaciones más groseras como DCT o Fourier local). El vector $\mathbf{c}$ es la imagen transformada, R es una transformación que puede ser lineal o no lineal y el vector $\mathbf{r}$ es la respuesta final a la imagen.

El estudio de estas transformaciones se puede abordar desde diferentes perspectivas:

- Psicofísica [11, 45, 46]: realizando experimentos que determinen los umbrales de discriminación (punto en el que un estímulo puede ser detectado, o puede ser detectada una variación respecto a un estímulo dado).
- Neurofisiolófica [2, 47, 14]: realizando medidas de respuesta en neuronas individuales o grupos de neuronas ante estímulos naturales.
- Estadística [48, 24, 49, 21]: a partir de muestras de $\mathbf{i}$ y alguna hipótesis sobre cuál es el objetivo que persigue el sistema se intenta obtener $\mathbf{W}$ o reproducir las propiedades de R. El uso de esta hipótesis ha permitido reproducir las propiedades de los campos receptivos y ha permitido reproducir propiedades de las no-linealidades utilizando un modelo concreto de no-linealidad.

Según esta última aproximación, la obtención de los modelos es importante ya que representan la adaptación, mediante evolución, del sistema visual humano a los estímulos a los que está sometido, y por lo tanto podrían ser utilizados como alternativa a los modelos basados en la estadística de la señal. Como veremos, estos modelos de respuesta se pueden utilizar para diseñar cuantizadores (la idea se puede expresar de una forma simple: colocar más bits donde es sistema visual es más sensible) y para definir operadores de penalización en el contexto de la restauración (la idea de nuevo se puede expresar de forma simple: no se debe penalizar en la solución la presencia de características que el sistema visual considera relevantes).

2.2. Métrica inducida por el modelo de respuesta

En esta sección mostraremos cómo un modelo de respuesta induce una métrica en cada uno de los diferentes dominios. El conocimiento de la métrica es importante ya que, como mostraremos en los capítulos siguientes, permite diseñar cuantizadores perceptuales, permite medir la similitud entre fotogramas y permite conocer cuales son las direcciones más y menos sensibles desde el punto de vista perceptual cuando estamos en presencia de un determinado estímulo.

En el dominio r se asume que la distancia entre un punto $\mathbf{r}$ y un punto cercano $\mathbf{r} + \Delta\mathbf{r}$ es Euclidea, es decir, el lugar geométrico de los puntos que están a una distancia $\Delta\mathbf{r}$ de un punto dado es una hiper-esfera, y además estas hiper-esferas son iguales en todos los puntos del dominio r,

$$d(\mathbf{r}, \mathbf{r} + \Delta\mathbf{r}) = \Delta\mathbf{r}^T \Delta\mathbf{r}$$

Puesto que la distancia debe ser igual independientemente del dominio y utilizando $\Delta\mathbf{r} = R(\mathbf{c} + \Delta\mathbf{c}) - R(\mathbf{c})$ tenemos

$$d(\mathbf{c}, \mathbf{c} + \Delta\mathbf{c}) = d(\mathbf{r}, \mathbf{r} + \Delta\mathbf{r}) = \big(R(\mathbf{c} + \Delta\mathbf{c}) - R(\mathbf{c})\big)^T \big(R(\mathbf{c} + \Delta\mathbf{c}) - R(\mathbf{c})\big) \quad (2.2)$$

El primer término del desarrollo en serie de Taylor de R en torno al punto $\mathbf{c}$ viene dado por

$$R(\mathbf{c} + \Delta\mathbf{c}) = R(\mathbf{c}) + \nabla R(\mathbf{c})\Delta\mathbf{c}$$

Sustituyendo esta expresión en la ecuación 2.2

$$d(\mathbf{c}, \mathbf{c} + \Delta\mathbf{c}) = \Delta\mathbf{c}^T \nabla R^T \nabla R \Delta\mathbf{c}$$

Por lo tanto, la matriz métrica, $\mathbf{G}$, en el dominio c viene dada por

$$\mathbf{G}(\mathbf{a}) = \nabla R(\mathbf{c})^T \nabla R(\mathbf{c})$$

$\nabla R(\mathbf{c})$ representa la sensibilidad del sistema en el punto $\mathbf{c}$: un $\nabla R(\mathbf{c})$ grande significa que una pequeña variación en la entrada $\Delta\mathbf{c}$ implica un cambio grande en la respuesta, y por lo tanto la métrica no es más que el cuadrado de la sensibilidad. Como se puede observar la elección de un modelo de respuesta induce una métrica en el dominio c (y también en el dominio i).

2.3. Modelo de respuesta lineal

En este caso, la respuesta r_f se define como [11]

$$r_f = R_f(c_f) = \mathrm{CSF}_f \cdot c_f \quad (2.3)$$

donde CSF_f es el valor de la función de sensibilidad al contraste[2] para la frecuencia f. La figura 2.1 muestra la forma de esta función y la respuesta para diferentes coeficientes. Como se puede observar para conseguir un mismo incremento de respuesta hay que aplicar un incremento de contraste diferente para cada coeficiente, sin embargo, para un coeficiente dado el incremento de contraste no depende del contraste.

La matriz métrica en este caso es una matriz diagonal y no depende del contraste del estímulo

[2] El contraste se define como la amplitud de un coeficiente dividido por la luminancia media.

$$G_{ff}(c_f) = \mathrm{CSF}_f^2$$

Con este modelo de respuesta el lugar geométrico de los puntos que están a una distancia $\mathbf{\Delta c}$ de un punto dado, c, es una hiper-elipse con los ejes orientados según las funciones base de $\mathbf{W}$. Además estas hiper-elipses son iguales en todos los puntos del dominio c.

Este modelo de respuesta es el que está en la base de JPEG [50, 41, 27] y se ha utilizado recientemente para definir perfiles adaptativos en el dominio DCT para máquinas de vectores soporte para la compresión de imágenes [51].

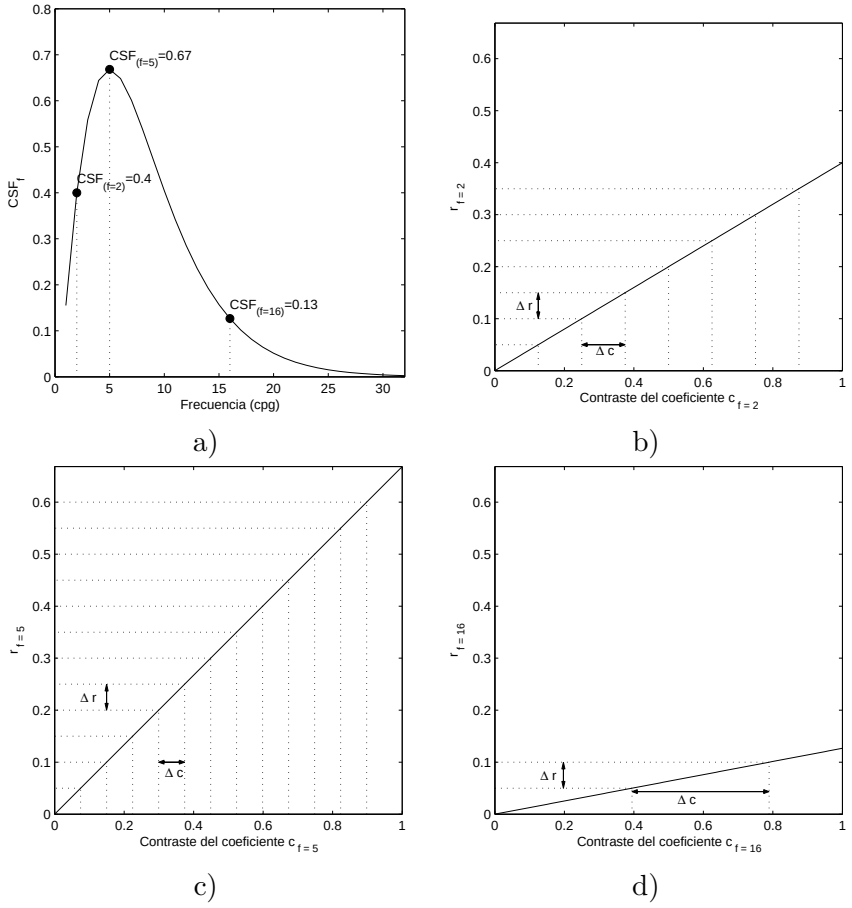


Fig. 2.1: Parámetro y ejemplos de respuesta para el modelo lineal. a) Forma de la función de sensibilidad al contraste. b), c) y d) ejemplos de respuesta para tres frecuencias. Los incrementos a aplicar en c para conseguir el mismo incremento en r dependen de la frecuencia pero no del contraste.

2.4. Modelo de respuesta no-lineal sin interacción

La respuesta, r_f, se define como [12, 13, 45]

$$r_f = R_f(c_f) = \frac{\alpha_f \cdot |c_f|^2}{\beta_f + |c_f|^2} \tag{2.4}$$

El ajuste de la respuesta a datos experimentales de umbrales incrementales proporciona los parámetros α_f y β_f que se muestran en la figura 2.2 junto con ejemplos de respuesta para tres frecuencias.

La matriz métrica en este caso es una matriz diagonal pero a diferencia del caso anterior sí que depende del contraste del estímulo,

$$G_{ff}(c_f) = c_f^2 \frac{4\alpha_f^2\beta_f^2}{(\beta_f + c_f^2)^4} \tag{2.5}$$

la figura 2.3 muestra la sensibilidad, $\nabla R_f(c_f)$, y la métrica $G_f(c_f)$ en función de la frecuencia y el contraste.

Este modelo de respuesta implica que el lugar geométrico de los puntos que están a una distancia $\mathbf{\Delta c}$ de un punto dado es una hiper-elipse orientada según las funciones base de $\mathbf{W}$ pero la longitud de los ejes de estas hiper-elipses crece con el contraste de $\mathbf{c}$.

Este modelo de respuesta siendo la etapa lineal una transformada de DCT local se utiliza en las publicaciones C, B y D.

2.5. Modelo de respuesta no-lineal con interacción

La respuesta, r_f, se define como [4, 5]

$$r_f(\mathbf{c}) = \frac{\alpha_f |c_f|^2}{\beta_f + \sum_{f'} \left(h_{ff'} |c_{f'}|^2\right)} \tag{2.6}$$

$h_{ff'}$ representa el peso que tiene cada coeficiente c'_f en la inhibición de la respuesta al coeficiente c_f. El peso es mayor cuanto más próximas son las frecuencias f y f'. Además la forma de este kernel de interacción depende de la transformada que se esté utilizando en la etapa lineal.

La existencia de estas relaciones entre coeficientes ha sido observada en diferentes dominios frecuenciales realizando estadística de coeficientes cuando los estímulos son imágenes naturales, por ejemplo en *wavelets* se ha mostrado que la

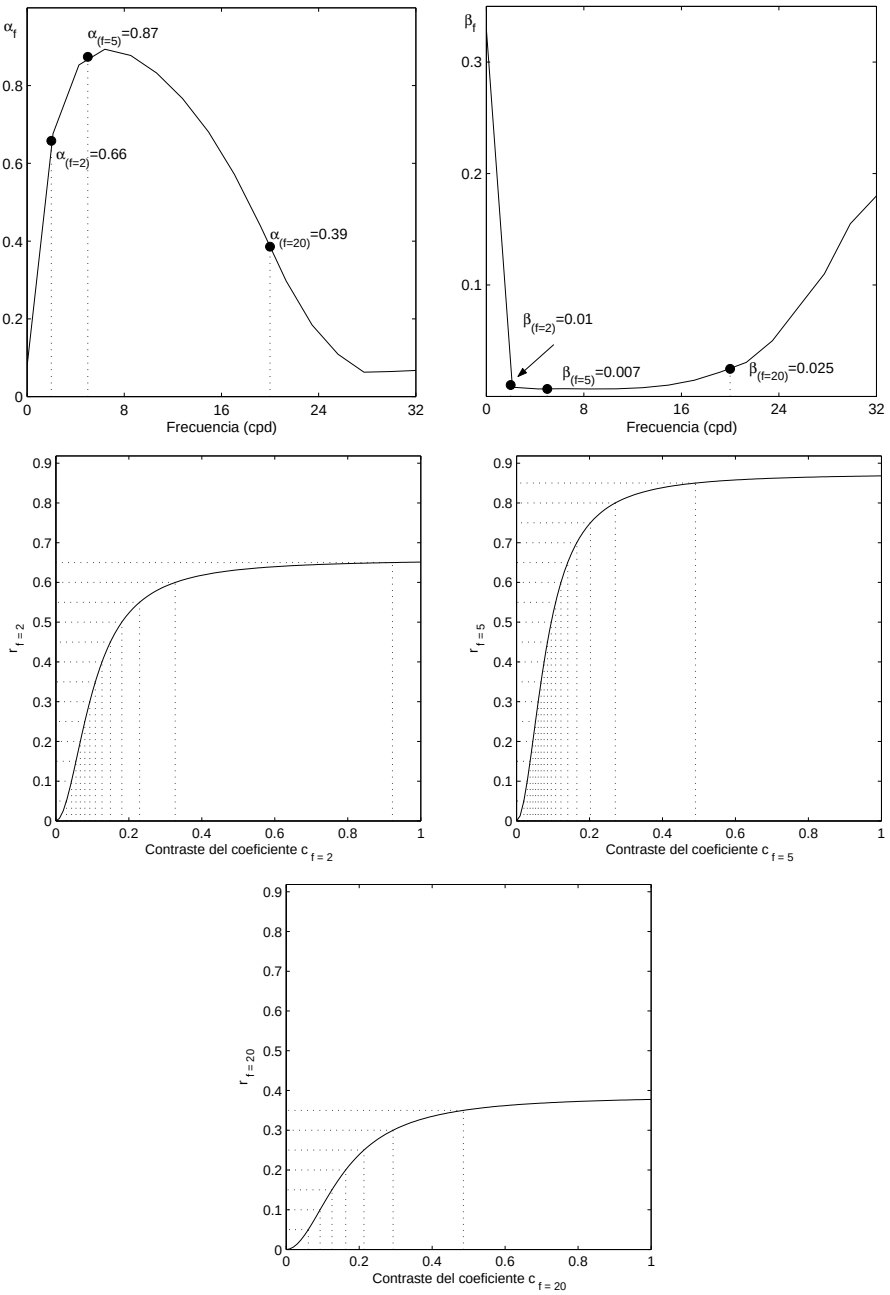


Fig. 2.2: Parámetros y ejemplos de respuesta para el modelo no lineal sin interacción. En la primera fila se muestra el ajuste de los parámetros α_f y β_f de la ecuación 2.4. En la segunda y tercera filas se muestran ejemplos de respuesta para tres frecuencias. En este caso, para conseguir incrementos iguales en r_f hay que aplicar incrementos que son dependientes del contraste.

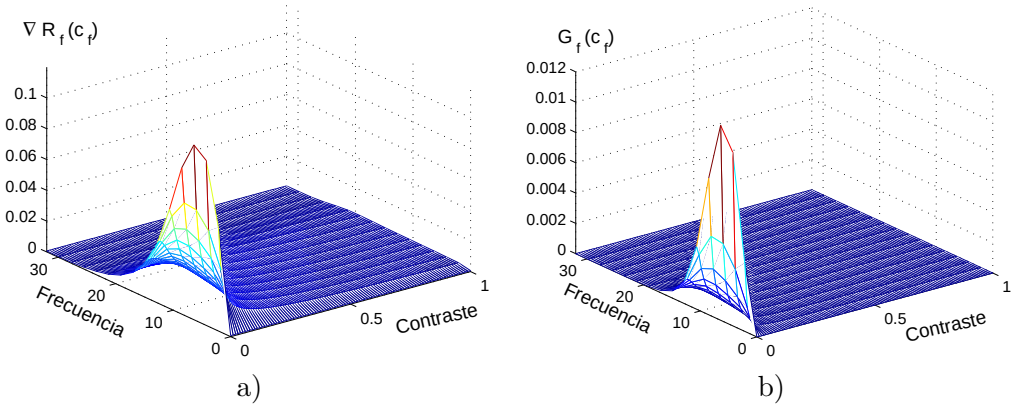


Fig. 2.3: Gradiente de la respuesta, a), y métrica, b), en función de la frecuencia y el contraste del estímulo para el modelo de respuesta no lineal sin interacción entre coeficientes.

probabilidad condicional de un coeficiente dado otro tiene una forma de pajarita [24, 26], indicando que si aumenta el valor del coeficiente dado, la varianza del otro coeficiente aumenta. Un efecto similar se ha mostrado en el dominio ICA [52]. En el dominio DCT local también se ha mostrado esta dependencia [53]. En la figura 1 del **Trabajo E** se muestra que estas relaciones también se dan en el dominio de Fourier local. Se puede observar que el número de coeficientes relacionados con uno dado se incrementa con la frecuencia del coeficiente.

Los parámetros que aparecen en esta ecuación y diversos ejemplos de respuesta se muestran en la figura 2.4 cuando la etapa lineal es una DCT local.

Utilizando este modelo, podemos ver cual es la variación de la respuesta cuando el estímulo está formado por dos funciones base. Una de las funciones base actúa como enmascaramiento y la otra función base como estímulo. La figura 2.5 muestra cual es la respuesta ante el estímulo (una función base DCT de frecuencia 10.6 cpg) para diferentes valores del contraste del enmascaramiento (una función base DCT de frecuencia 15 cpg) cuando se utiliza la ecuación 2.6.

Para este modelo de respuesta, la matriz métrica es no diagonal y depende de la entrada:

$$G_{ff'}(\mathbf{c}) = 4\alpha_f^2 \left(\frac{c_f}{\beta_f + \sum_{f'} \left(h_{ff'} |c_{f'}|^2 \right)} \delta_{ff'} - \frac{c_f^2 c_{f'}}{\left(\beta_f + \sum_{f'} (h_{ff'} |c_{f'}|^2) \right)^2} h_{ff'} \right)^2$$

Según este modelo de respuesta el lugar geométrico de los puntos que están a una distancia $\Delta\mathbf{c}$ de un punto dado es una hiper-elipse que no está orientada

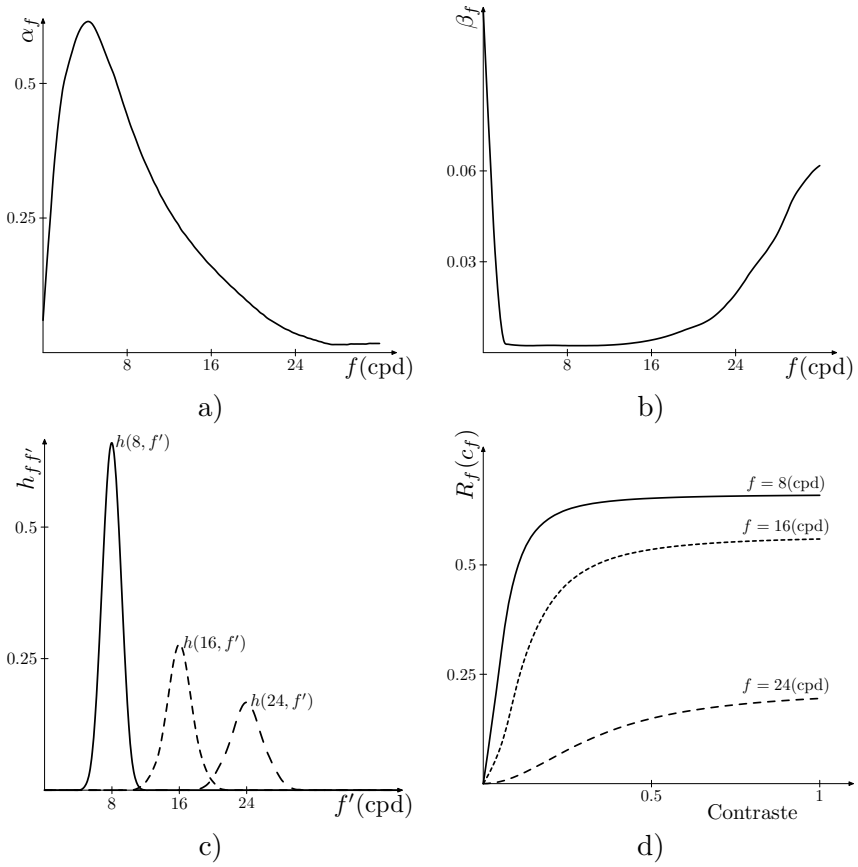


Fig. 2.4: Parámetros y ejemplos de respuesta para el modelo no lineal con interacción. a) α_f, b) β_f, c) $h_{ff'}$ y d) ejemplos de respuesta para tres frecuencias. Para obtener estas respuestas se ha considerado que el estímulo contiene únicamente una función base de la frecuencia indicada.

según las funciones base de $\mathbf{W}$ y además la longitud de los ejes de estas hiperelipses depende del estímulo $\mathbf{c}$.

Este modelo de respuesta siendo la etapa lineal una transformada de Fourier local se utiliza en el Trabajo E.

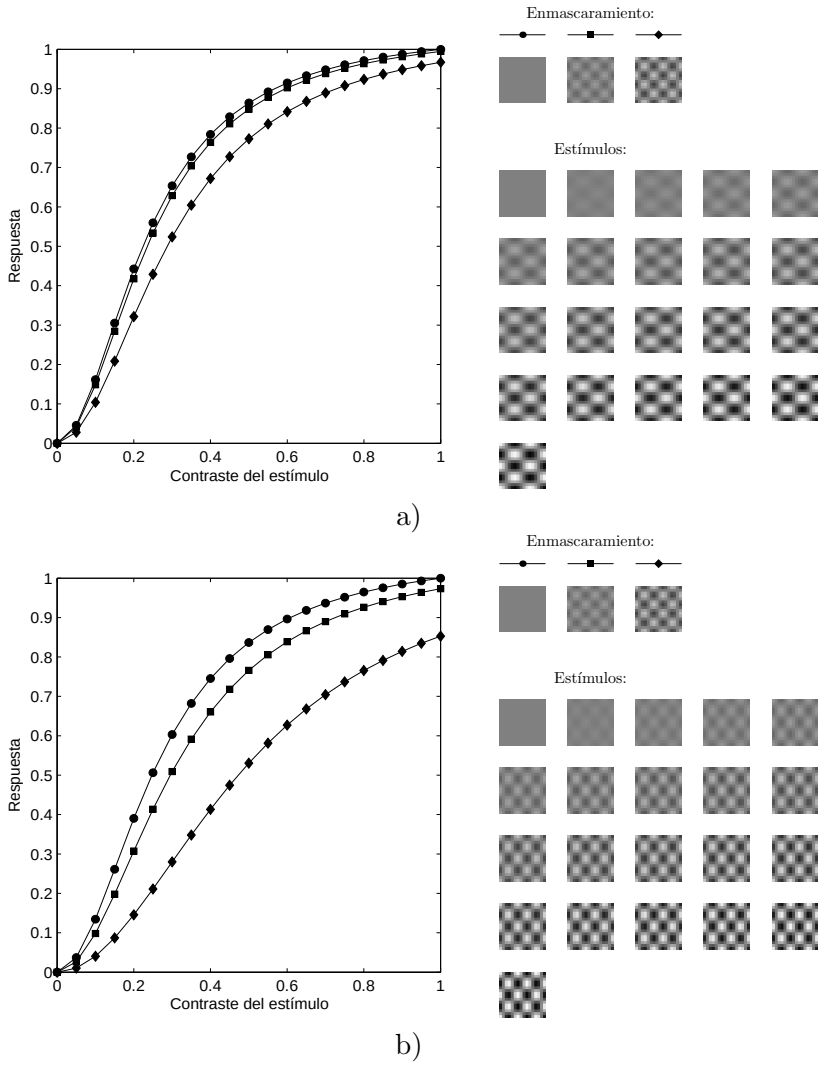


Fig. 2.5: Respuesta ante dos estímulos para diferentes contrastes del enmascaramiento calculada utilizando la ecuación 2.6. a) El estímulo es una función base DCT de frecuencia 10.6 cpg y el enmascaramiento es otra función base DCT de frecuencia 15 cpg. b) El estímulo es una función base DCT de frecuencia 13.6 cpg y el enmascaramiento es el mismo que el anterior. Como se puede observar, en ambos casos, cuando se incrementa el contraste del enmascaramiento (círculo: contraste 0, cuadrado: contraste 0.2 y rombo: contraste 0.5) la respuesta ante el estímulo disminuye. Sin embargo esta disminución es más acusada si las frecuencias del estímulo y del enmascarador están más próximas.

3. OBTENCIÓN DE PROPIEDADES DE LAS NO LINEALIDADES DE V1 A PARTIR DEL ANÁLISIS EN COMPONENTES INDEPENDIENTES LOCAL

En los capítulos anteriores se ha comentado que las propiedades de la representación de imágenes en las etapas de bajo nivel del sistema visual humano se pueden entender mediante el estudio de la estadística de los estímulos naturales.

A partir de una serie de estímulos naturales y de alguna hipótesis que indique cuál es el objetivo que se persigue con la transformación o secuencia de transformaciones es posible reproducir propiedades que se han obtenido mediante experimentos psicofísicos o neurofisiológicos. La siguiente lista muestra tres ejemplos representativos:

- Por un lado Joseph J. Atick [22] explora el principio de la representación eficiente de la información como un principio que guía el diseño del proceso sensorial, de este modo muestra cómo utilizando este principio se puede obtener la función de sensibilidad al contraste (CSF).

- Por otro lado, B. A. Olshausen y D. J. Field [20] muestran que un algoritmo que intenta encontrar códigos lineales dispersos (se busca un código de baja entropía en el cual la distribución de probabilidad de actividad de cada coeficiente sea unimodal y con un pico en cero) para imágenes naturales desarrolla una familia completa de campos receptivos localizados, pasabanda y orientados, similares a los encontrados en el cortex primario (V1). Es decir, reproducen propiedades de la etapa lineal.

- Finalmente, O. Schwartz and E. Simoncelli [24] obtienen los parámetros del modelo normalización divisiva, asumiendo que las respuestas deben ser independientes. Muestran que esto proporciona una buena caracterización de las propiedades de respuesta no lineal de neuronas típicas en el cortex visual primario o en el nervio auditivo.

El trabajo presentado en este capítulo también se centra en la reproducción del comportamiento no lineal de V1 utilizando la hipótesis de la codificación eficiente. No obstante, su originalidad radica en el hecho de que a diferencia de [24], no se

impone ninguna forma funcional particular para la no linealidad de manera que se resalta el hecho de que ésta emerge directamente de los datos.

Desde el punto de vista técnico, en este trabajo la transformación no lineal surge de la aplicación de una solución novedosa del Análisis en Componentes Independientes (ICA) no lineal.

3.1. ICA lineal y no lineal

Un problema central en diversas disciplinas (estadística, procesado de señales, análisis de datos,...) es la búsqueda de una representación adecuada de los datos. Es importante que los datos estén representados de tal forma que se facilite su análisis posterior, ya se trate de reconocimiento de patrones, de compresión de los datos, de eliminación de ruido, etc. En determinadas ocasiones el objetivo de esta transformación es conseguir que los datos en esta nueva transformación sean estadísticamente independientes.

En la literatura se pueden encontrar diferentes principios y métodos desarrollados para obtener transformaciones lineales o no lineales apropiadas para conseguir el anterior objetivo. Algunos ejemplos son el análisis en componentes principales (PCA) o transformada de Karhunen-Loeve y el análisis en componentes independientes. Estos métodos definen un principio que especifica bajo qué criterios la transformación es óptima. Por ejemplo en el PCA la transformación lineal óptima es aquella que elimina las correlaciones de segundo orden de los datos transformados, sin embargo este criterio no garantiza que los datos sean independientes en esta representación cuando la PDF no es Gaussiana.

En el análisis en componentes independientes el objetivo básico es encontrar una transformación de los datos de forma que se maximice la independencia entre las componentes s_i s.

En el caso del ICA lineal [44, 54] el problema se puede formular como la estimación del siguiente modelo generativo de los datos, donde los datos i se obtienen a partir de una combinación lineal de un conjunto de fuentes, r_j, independientes,

$$\mathbf{i} = \mathbf{A}\mathbf{r} \tag{3.1}$$

donde la matriz $\mathbf{A}$ se conoce como matriz de mezcla. La matriz $\mathbf{W}$ (que coincidirá con $\mathbf{A}^{-1}$ en el caso de matrices cuadradas) que lleva del dominio i al dominio r se conoce como matriz de separación.

Este modelo es, en general, insuficiente ya que por una parte es lineal, y por otra parte describe todos los datos utilizando las mismas características globales (mediante un único conjunto de funciones base).

A pesar de estas debilidades hay que decir que uno de los éxitos del ICA cuando se aplica sobre un conjunto de imágenes naturales es que se obtienen funciones base con propiedades similares a los campos receptivos (es decir, localización espacial, filtros pasa banda y orientados) [21, 55, 20] a diferencia de lo que se obtendría utilizando un análisis en componentes principales (que proporcionan un conjunto de funciones base tipo DCT). La figura 3.1 muestra un ejemplo de las funciones base que se obtendrían utilizando como datos de entrenamiento estímulos correspondientes de imágenes naturales de tamaño 16×16.

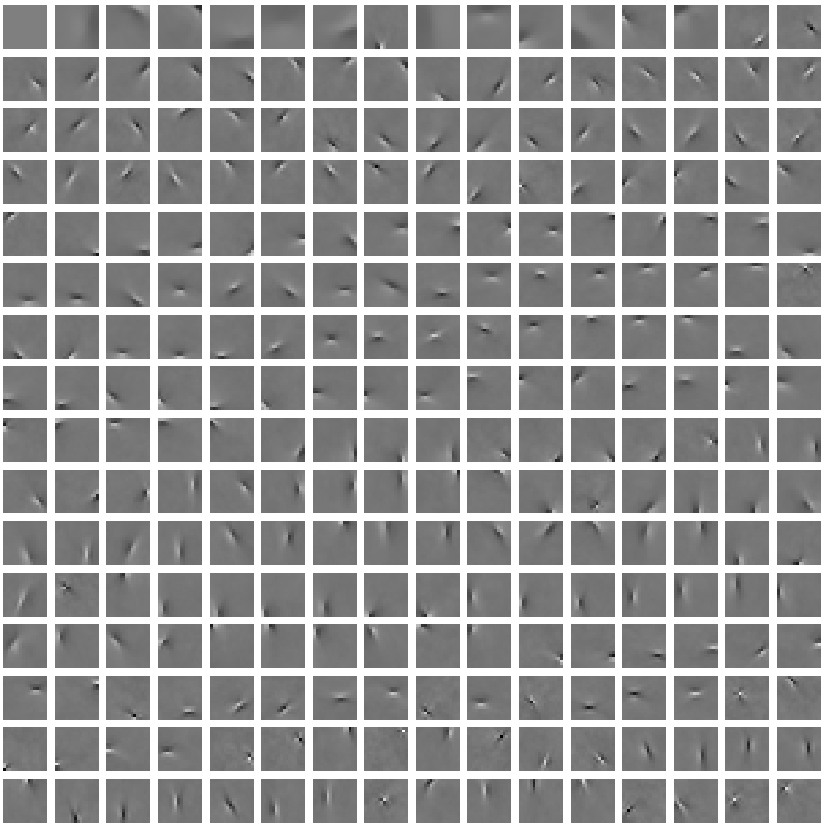


Fig. 3.1: Columnas de la matriz de mezcla A. Los datos de entrada para la obtención de esta matriz mezcla son estímulos naturales de tamaño 16×16.

El análisis en componentes independientes no lineal extiende el modelo anterior asumiendo que las componentes de los vectores de datos $\mathbf{i}$ dependen de un modo no lineal de las componentes independientes $\mathbf{r}$,

$$\mathbf{i} = F(\mathbf{r}) \tag{3.2}$$

donde F es una función no lineal de mezcla. El problema, al igual que en el caso

anterior, consiste en la estimación de las componentes independientes **s** utilizando únicamente observaciones de las mezclas **i**. Sin embargo en el caso no lineal no existe una solución única [56] siendo necesario imponer algún tipo de restricción.

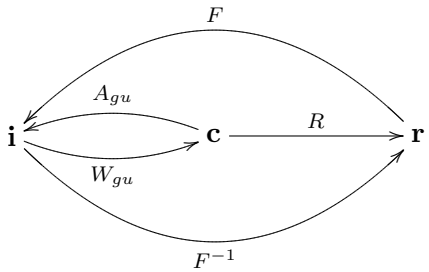
Los métodos que se basan en métodos lineales [44, 29, 30] dividen los datos en agrupaciones y obtienen un modelo lineal en cada una de ellas. Mediante las agrupaciones se consigue dividir el problema no lineal global en una serie de subproblemas en los que se puede adoptar una aproximación lineal local.

Sin embargo, tal y como vamos a mostrar, es posible relacionar los subproblemas locales lineales para obtener una representación global no lineal.

3.2. No linealidades de la percepción a partir de un ICA no lineal

Si el objetivo de la representación cortical, según la hipótesis de la codificación eficiente, es conseguir la independencia de los coeficientes en esta representación, podemos considerar que la secuencia de transformaciones mostrada en la ecuación (2.1) del capítulo 2 se puede contemplar como un ICA no lineal.

En este esquema, hay una parte que asumimos conocida (la parte lineal), y otra parte (la no lineal) que pretendemos obtener a partir de los datos.



(3.3)

Esta secuencia de transformaciones se ilustra en la figura 3.2.

En cuanto a la parte lineal, las funciones campo receptivo se modelan mediante las filas de la matriz de separación $\mathbf{W}_{gu}$ (siguiendo las ideas de Olshausen *et al.* [20]). Esta matriz se ha obtenido realizando un ICA lineal [57] global. Las filas de esta matriz han sido normalizadas ya que habitualmente se asume que esta matriz es unitaria.

A partir de una solución de F podremos obtener las propiedades de la transformación no lineal R.

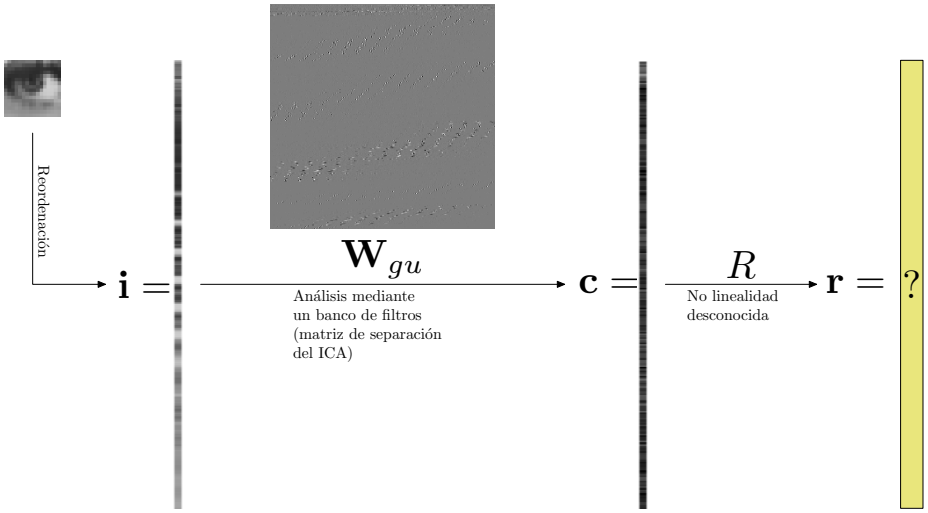


Fig. 3.2: Ilustración de la secuencia de transformaciones que sufre un estímulo. Primero el estímulo (un bloque de una imagen) se reordena para obtener un vector. A continuación se analiza mediante la matriz de separación global unitaria $\mathbf{W}_{gu}$. Finalmente se aplica una no linealidad R desconocida.

3.3. Solución del ICA no lineal mediante una aproximación diferencial

Si se asume que la aplicación no lineal 3.2 es diferenciable, entonces puede ser aproximada localmente [29] por el modelo de mezcla lineal (3.1). Es decir

$$\nabla F(\mathbf{r_0}) = \mathbf{A}\big(F(\mathbf{r}_0)\big) \tag{3.4}$$

donde $\mathbf{A}\big(F(\mathbf{r}_0)\big)$ es la matriz de mezcla obtenida utilizando estímulos seleccionados en torno al punto indicado. Basándonos en este resultado y en el modelo presentado en la sección anterior, vamos a proponer una solución particular del ICA no lineal mediante una aproximación diferencial. Según la ecuación 3.3

$$\mathbf{i} = F(\mathbf{r}) = \mathbf{A}_{gu} R^{-1}(\mathbf{r})$$

de donde se deduce derivando que

$$\nabla F(\mathbf{r_0}) = \mathbf{A}_{gu} \nabla R^{-1}(\mathbf{r_0}) = \mathbf{A}\big(F(\mathbf{r}_0)\big) \tag{3.5}$$

invirtiendo se obtiene

$$\nabla R(\mathbf{c_0}) = \mathbf{W}(\mathbf{i}_0)\mathbf{A}_{gu} \tag{3.6}$$

La respuesta ante un estímulo **i**, equivalentemente **c**, se puede calcular como la suma de las variaciones locales de la respuesta

$$\mathbf{r} = R(\mathbf{c}) = \mathbf{r}(\mathbf{0}) + \int_{\mathbf{0}}^{\mathbf{c}} \nabla R(\mathbf{c}')d\mathbf{c}' \tag{3.7}$$

sustituyendo el resultado obtenido en la ecuación (3.6) se obtiene que la respuesta, **r**, ante cualquier estímulo, **i**, se puede obtener como

$$\mathbf{r} = \int_{E\{\mathbf{I}\}}^{\mathbf{i}} \mathbf{W}(\mathbf{i}')d\mathbf{i}' \tag{3.8}$$

donde $E\{\mathbf{I}\}$ es el valor esperado de los estímulos (que para los datos considerados se corresponde a un estímulo de contraste 0 para el cual la respuesta es $\mathbf{0}$), y la integración se debe realizar numéricamente (por ejemplo utilizando el método de Runge-Kutta [58]).

Es decir, para calcular la respuesta hay que obtener la matriz de separación $W(\mathbf{i}')$ en una serie de puntos de una ruta que una el punto de respuesta nula ($E\{\mathbf{I}\}$) y el punto para el que se desea obtener la respuesta **i**.

Hay que resaltar dos aspectos importantes de esta formulación. El primero de ellos es que los coeficientes de la respuesta del ICA no lineal deberían ser más independientes que los coeficientes del ICA lineal global. En segundo lugar la integral de la ecuación (3.8) debería ser independiente del camino de integración. Estos dos aspectos son analizados en la sección de resultados.

3.4. Dirección (imagen) que excita a un único sensor

En esta sección vamos a buscar la expresión que nos permita obtener la variación que hay que aplicar sobre un estímulo para que esta variación provoque un incremento de respuesta igual a τ en un único coeficiente de la respuesta (en un único sensor). Esta idea se ilustra en la figura 3.3.

Este resultado nos servirá para definir las rutas de integración en la ecuación 3.8 para obtener curvas de respuesta para cada sensor de forma aislada y las curva de respuesta en presencia de un estímulo que actúa de enmascaramiento.

Según el modelo de transformaciones propuesto en la ecuación (3.3) la repuesta, **r**, ante un estímulo, **i**, es

$$\mathbf{r} = R(\mathbf{W}_{gu}\mathbf{i})$$

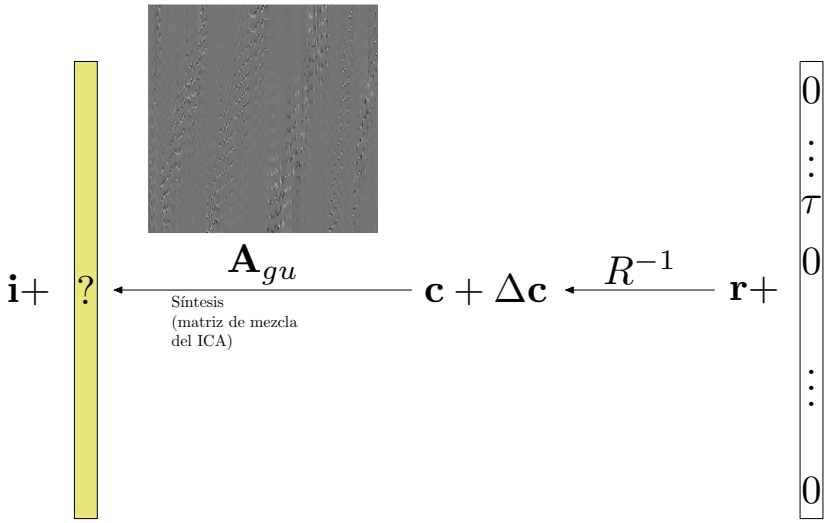


Fig. 3.3: Búsqueda del estímulo que excita a un único sensor

Si se presenta una ligera modificación del estímulo, $\mathbf{i} + \Delta\mathbf{i}$, y se aproxima el valor de la respuesta utilizando el primer término del desarrollo en serie se tiene que,

$$\mathbf{r} + \Delta\mathbf{r} \approx R(\mathbf{c}) + \nabla R(\mathbf{c})\Delta\mathbf{c} = R(\mathbf{c}) + \nabla R(\mathbf{c})\mathbf{W}_{gu}\Delta\mathbf{i}$$

de donde se deduce que el incremento de respuesta, $\Delta\mathbf{r}$, provocado por la diferencia del estímulo, $\Delta\mathbf{i}$, es

$$\Delta\mathbf{r} = \nabla R(\mathbf{c})\mathbf{W}_{gu}\Delta\mathbf{i}$$

Puesto que buscamos el estímulo que excita al sensor k una cantidad τ, sustituimos $\Delta\mathbf{r}$ por $\tau\boldsymbol{\delta}^k$ ($\boldsymbol{\delta}^k$ es un vector con un uno en la posición k y ceros en el resto de posiciones) en la expresión anterior,

$$\tau\boldsymbol{\delta}^k = \nabla R(\mathbf{c})\mathbf{W}_{gu}\Delta\mathbf{i}$$

y despejando $\Delta\mathbf{i}$ se llega a

$$\Delta\mathbf{i} = \mathbf{A}_{gu}\nabla R(\mathbf{c})^{-1}\tau\boldsymbol{\delta}^k \tag{3.9}$$

Esta ecuación se podría utilizar en el caso en el que disponga de un modelo para R, este es el caso presentado en [59] en el que se utilizaba el modelo de respuesta no lineal con interacción entre coeficientes dado por la ecuación (2.6).

En el caso que nos ocupa no disponemos de modelo para R, sin embargo, podemos sustituir en esta expresión el resultado de la ecuación (3.6) obteniendo

$$\Delta\mathbf{i} = \mathbf{A}_{gu}\mathbf{A}_{gu}^{-1}\mathbf{W}(\mathbf{i})^{-1}\tau\boldsymbol{\delta}^k$$

simplificando se llega a la expresión que nos da la dirección que estimula al sensor k,

$$\Delta\mathbf{i} = \tau\mathbf{A}(\mathbf{i})\boldsymbol{\delta}^k \tag{3.10}$$

Es decir, la dirección que hace que se estimule el sensor k a partir de un estímulo $\mathbf{i}$ es la dirección dada por la columna k de la matriz mezcla del ICA local en torno al estímulo $\mathbf{i}$.

Sustituyendo la ecuación (3.10) en la ecuación (3.8) y teniendo en cuenta la definición del contraste de Michelson [60] de los estímulos incrementales,

$$C(\Delta\mathbf{i}) = \frac{\text{máx}(\Delta\mathbf{i}) - \text{mín}(\Delta\mathbf{i})}{2L_0}$$

donde L_0 es la luminancia media de $E\{\mathbf{I}\}$, podremos obtener curvas de respuesta frente al contraste cuyas propiedades vamos a analizar.

3.5. Experimentos y resultados

3.5.1. Hechos psicofísicos a reproducir

En la sección anterior se ha mostrado cómo se puede obtener la respuesta ante un estímulo utilizando únicamente la distribución de estímulos naturales. En la introducción a este capítulo se comentó que el objetivo del estudio de la estadística de las imágenes naturales es la obtención de conocimiento sobre el funcionamiento del sistema visual. Por lo tanto, en esta sección nos planteamos comprobar qué propiedades psicofísicas se pueden reproducir con este modelo.

Las propiedades de la etapa lineal (el hecho de que sean filtros pasa banda orientados) están impuestas, ya que en la primera etapa se realiza un análisis del estímulo utilizando la matriz de separación $\mathbf{W}_{gu}$ del ICA. Las filas de esta matriz contienen filtros pasa banda orientados.

Puesto que en este trabajo nos centramos en la parte no lineal, sí que debemos comprobar si se cumplen las siguientes propiedades de las no linealidades de la respuesta. Concretamente:

- **Saturación de la respuesta**: cuando un determinado sensor se excita en presencia de su estímulo óptimo la respuesta se satura.

- **Efecto del contraste en el enmascaramiento cruzado**: la respuesta ante el estímulo óptimo decrece cuando se está en presencia de otro estímulo que actúa de enmascaramiento. El decremento en la respuesta crece con el contraste del estímulo de enmascaramiento.

- **Efecto de la frecuencia en el enmascaramiento cruzado**: cuanto más próxima sea la frecuencia del enmascaramiento a la del estímulo mayor será el decremento en la respuesta.

Estas tres propiedades se pueden observar en la figura 2.5 del capítulo 2 en la que se ha utilizado una expresión concreta para la no linealidad, R.

3.5.2. Detalles de la implementación

Los resultados que se presentan en esta sección se han realizado utilizando $2 \cdot 10^6$ estímulos de tamaño 6×1 pixels extraídos de forma aleatoria de la base de imágenes naturales de van Hateren [55]. Para calcular cada uno de los ICAs locales se han tomado los $5 \cdot 10^4$ estímulos más cercanos al punto considerado (usando distancia Euclídea en el dominio i). Este número de vecinos se ha elegido para que se tengan suficientes estímulos para calcular los ICAs locales y a la vez que se cumpla que este número no sea excesivamente grande de forma que capture las propiedades locales (conforme más puntos se tomen, más parecido será el resultado del ICA local al del ICA global).

El hecho de no utilizar estímulos de mayor tamaño, $N \times N$, viene determinado por el coste computacional del proceso descrito en la sección 3.3. El cuello de botella se produce en la obtención del ICA local [57] (sobre un espacio de N^2 dimensiones) ya que este proceso se debe llevar a cabo en múltiples ocasiones. Este problema computacional se podría solventar utilizando un PCA local [30].

La figura 3.4 muestra la distribución de los puntos en el dominio imagen para diferentes planos y la figura 3.6 muestra la distribución de los puntos una vez transformados al dominio ICA global (mediante la matriz $\mathbf{W}_{gu}$).

Realizando un ICA lineal global sobre estos estímulos se obtiene la siguiente matriz de mezcla

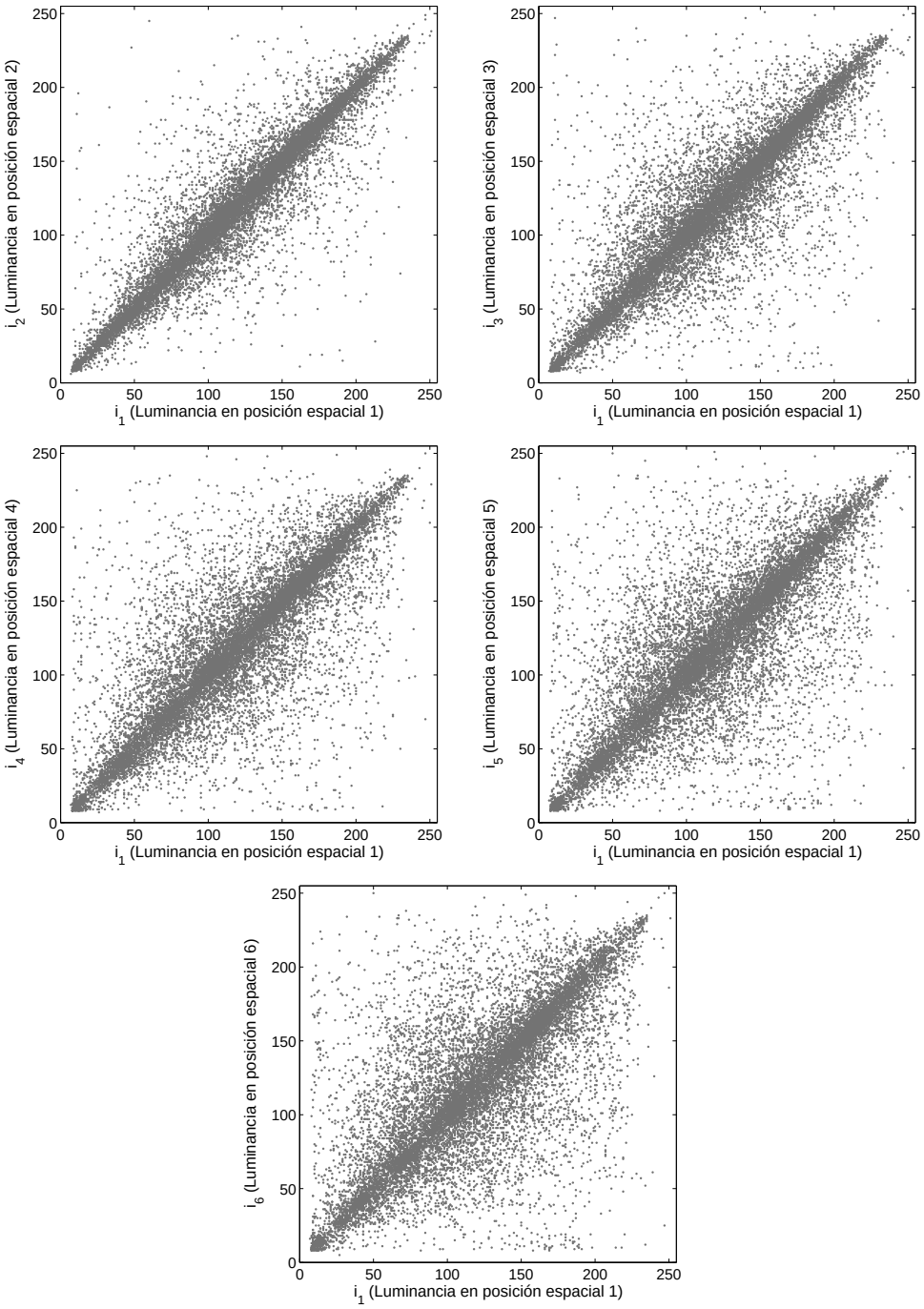


Fig. 3.4: Valores de luminancia en puntos vecinos del dominio espacial para diferentes pares de posiciones.

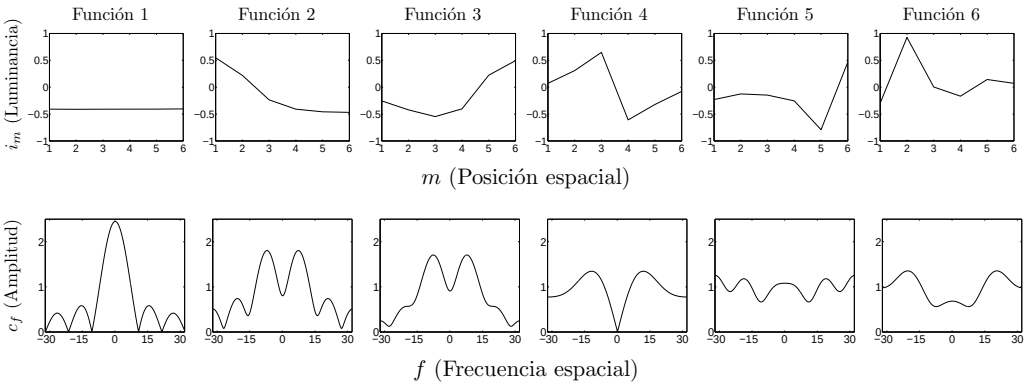


Fig. 3.5: Columnas de la matriz mezcla, $\mathbf{A}_{gu}$, del ICA global en el dominio espacial y en el dominio frecuencial.

$$\mathbf{A}_{gu} = \begin{bmatrix} -0{,}4084 & 0{,}5477 & -0{,}2552 & 0{,}0717 & -0{,}2300 & -0{,}2956 \\ -0{,}4113 & 0{,}2202 & -0{,}4223 & 0{,}3091 & -0{,}1251 & 0{,}9269 \\ -0{,}4094 & -0{,}2354 & -0{,}5471 & 0{,}6477 & -0{,}1469 & 0{,}0041 \\ -0{,}4085 & -0{,}4087 & -0{,}4026 & -0{,}6084 & -0{,}2551 & -0{,}1669 \\ -0{,}4076 & -0{,}4579 & 0{,}2239 & -0{,}3217 & -0{,}7911 & 0{,}1432 \\ -0{,}4042 & -0{,}4684 & 0{,}4951 & -0{,}0782 & 0{,}4679 & 0{,}0720 \end{bmatrix}$$

La figura 3.5 muestra las funciones base ICA de la matriz de mezcla global (columnas de la matriz $\mathbf{A}_{gu}$) junto con sus respectivos espectros, estas funciones se han ordenado según su norma (antes de realizar la normalización).

La figura 3.6 muestra la distribución de pares de coeficientes del dominio ICA global (correspondientes a la transformación mediante $\mathbf{W}_{gu}$ de los estímulos).

En algunos experimentos es necesario que el punto inicial de la integración sea el punto $E\{\mathbf{I}\}$, que para el caso particular de los estímulos considerados es el punto

$$E\{\mathbf{I}\} = \begin{bmatrix} 118{,}09 \\ 118{,}06 \\ 118{,}03 \\ 118{,}01 \\ 117{,}99 \\ 117{,}96 \end{bmatrix}$$

A modo de ejemplo, la figura 3.7 muestra las funciones base ICA de la matriz de mezcla local en torno al punto $E\{\mathbf{I}\}$ (tomando los $5 \cdot 10^4$ estímulos más próximos) siguiendo la misma ordenación que en el caso de la matriz global.

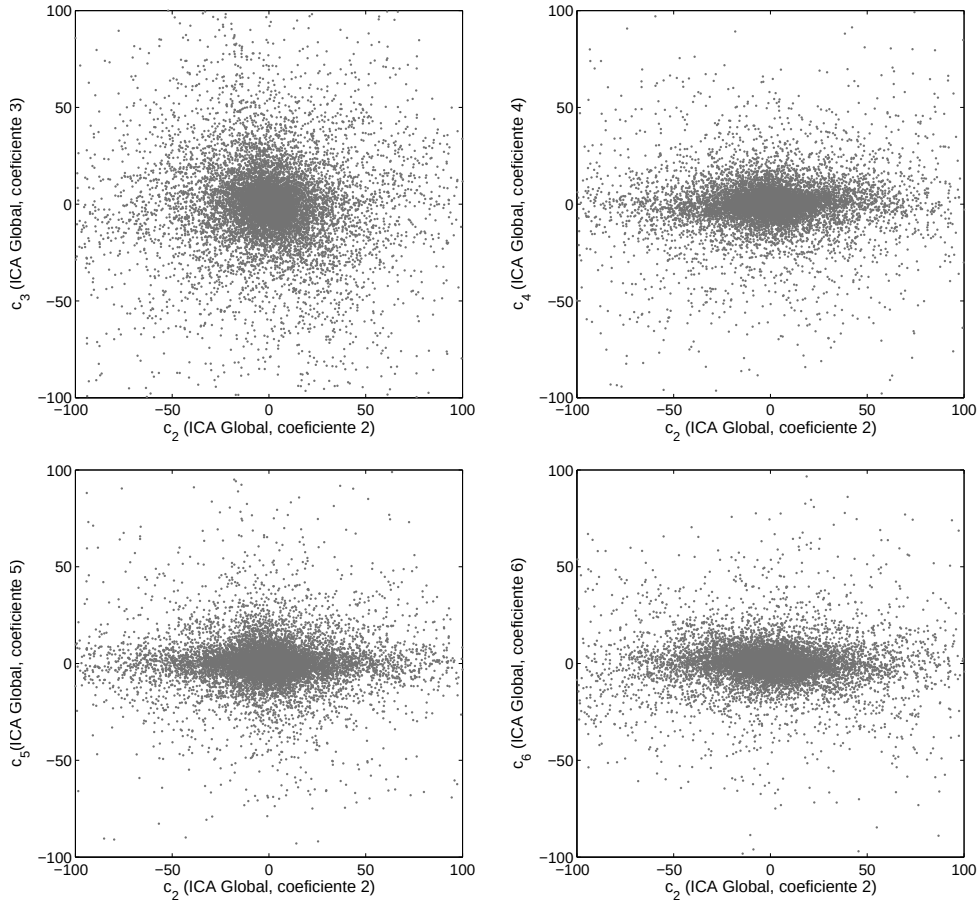


Fig. 3.6: Valores de coeficientes vecinos de la transformada ICA global.

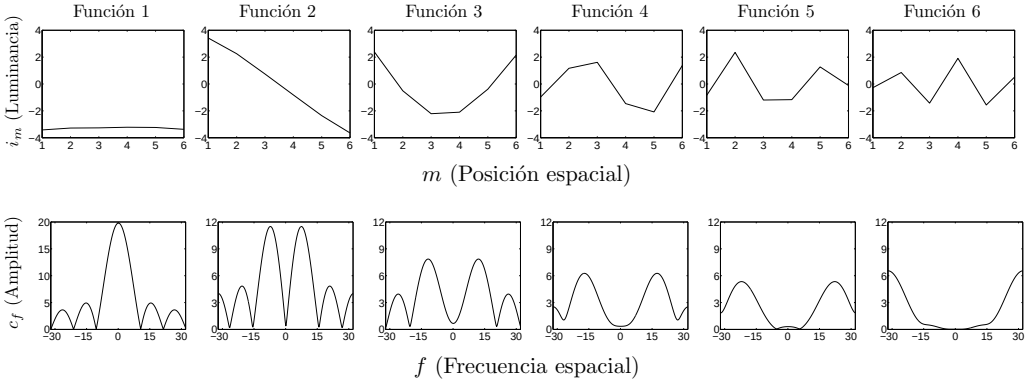


Fig. 3.7: Columnas de la matriz mezcla, A, del ICA local, en torno al punto $E\{\mathbf{I}\}$, en el dominio espacial y en el dominio frecuencial.

Un problema asociado al uso del ICA en cada punto de los trayectos es que no hay ordenación de las funciones base (columnas de la matriz $\mathbf{A}(\mathbf{i})$). Esto constituye un problema ya que se debe mantener un criterio a la hora de elegir el desplazamiento que excita a un determinado sensor. La distribución de los datos que estamos utilizando (estímulos naturales) hace posible que se pueda utilizar el criterio de ordenar las columnas de la matriz $\mathbf{A}(\mathbf{i})$ de forma que sea lo más parecida posible a la ordenación de las columnas de la matriz $\mathbf{A}(E\{\mathbf{I}\})$ En este caso, se ha utilizado un algoritmo voraz de asignación cuyo criterio es tomar la asignación que maximice el producto escalar de las columnas normalizadas de $\mathbf{A}(\mathbf{i})$ y de $\mathbf{A}(E\{\mathbf{I}\})$.

3.5.3. Saturación de la respuesta

Este experimento consiste en ver cómo evoluciona la respuesta frente al contraste del estímulo. Para ello, se parte del estímulo de contraste cero y nos vamos desplazando por la dirección que provoca que se estimule al sensor en cuya respuesta estamos interesados.

La figura 3.8 muestra cómo varían las componentes 2 y 4 de la columna 2 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 2) y de la columna 4 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 4) en el dominio ICA global. La longitud de los ejes representa la amplitud (o el contraste) de los estímulos incrementales. El hecho de que estas longitudes crezcan con el contraste es lo que da lugar a la saturación de la respuesta.

La figura 3.9 muestra las curvas de respuesta obtenidas para sensores de baja, media y alta frecuencia. Como se puede observar en esta figura, la respuesta se satura siendo este resultado consistente con experimentos psicofísicos y neurofisiológicos [12, 13, 4].

3.5.4. Enmascaramiento cruzado

En este caso partimos del estímulo con un determinado contraste que excita al sensor k y nos movemos por la ruta que excita al sensor j con el fin de comprobar cómo afecta la presencia de un estímulo de enmascaramiento, k, en la respuesta del sensor j.

La figura 3.10 muestra cómo varían las componentes 2 y 4 de la columna 2 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 2) y de la columna 4 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 4) en el dominio ICA global, para diferentes puntos de partida utilizando como enmascaramiento el estímulo que excita al sensor 2.

La figura 3.11 muestra las curvas de respuesta obtenidas para diferentes senso-

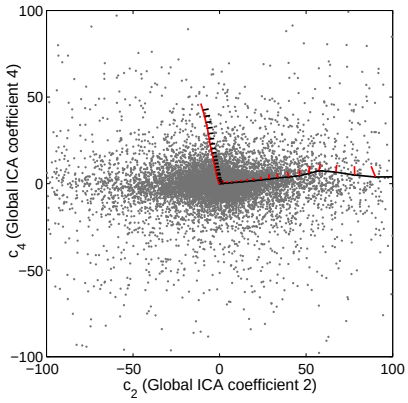


Fig. 3.8: Componentes 2 y 4 de la columna 2 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 2) y de la columna 4 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 4) en el dominio ICA global

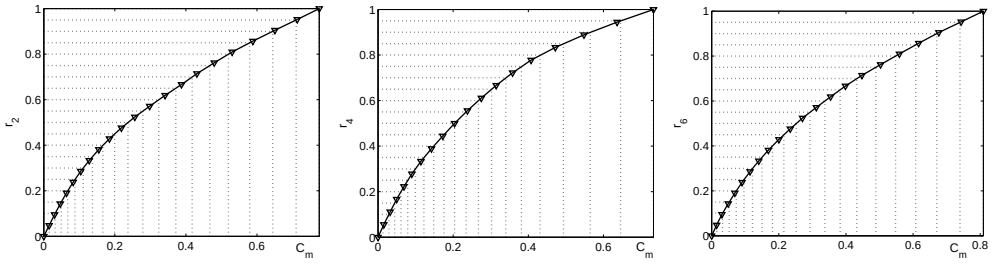


Fig. 3.9: Respuesta (normalizada al máximo en cada caso) frente al contraste para diferentes sensores, a) baja frecuencia, b) frecuencia media y c) alta frecuencia

res utilizando diferentes estímulos de enmascaramiento. Como se puede observar en esta figura, la respuesta disminuye en presencia de un estímulo de enmascaramiento y esta disminución es mayor cuanto más contraste tiene el enmascaramiento. Sin embargo no se reproduce la segunda propiedad asociada al enmascaramiento, es decir, estímulos de enmascaramiento de frecuencia más próxima no provocan una disminución mayor de la respuesta que la que provocan estímulos de enmascaramiento de frecuencia más alejada a la del estímulo.

3.5.5. Independencia de las componentes de la transformada local-a-global

Puesto que el modelo completo se asume que es un ICA no lineal cabe esperar que la independencia de sus componentes sea mayor que la independencia de las

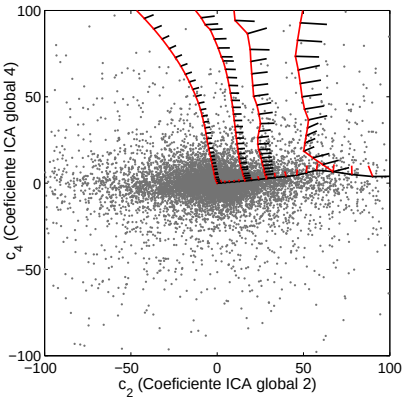


Fig. 3.10: componentes 2 y 4 de la columna 2 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 2) y de la columna 4 de la matriz $\mathbf{A}(\mathbf{i})$ (correspondiente a la ruta que excita al sensor 4) en el dominio ICA global, para diferentes puntos de partida utilizando como enmascaramiento el estímulo que excita al sensor 2.

componentes del ICA lineal global.

Para comprobar la independencia se han seleccionado aleatoriamente 10^4 estímulos,$\{\mathbf{i}\}$, del conjunto de $2{\cdot}10^6$ estímulos considerados, estos estímulos se han transformado mediante la matriz $\mathbf{W}_{gu}$ para obtener los coeficientes en el dominio ICA global, $\{\mathbf{c}\}$, así como las respuestas utilizando el procedimiento descrito en la sección 3.3, $\{\mathbf{r}\}$. En cada uno de estos dominios se han realizado dos experimentos, uno cualitativo en el que se muestra cual es la probabilidad condicional de un componente dado otro componente (los resultados de este análisis presentados en gráficamente dan lugar a una forma de pajarita, *bow tie*, cuando se estudia la relación entre coeficientes del dominio wavelet [24]) y otro cuantitativo en el que se mide la información mutua y la información mutua relativa de una componente dada otra.

La figura 3.12 muestra las probabilidades de pares de componentes en el domino espacial, en el dominio ICA lineal global y en el dominio propuesto.

En el dominio espacial se puede observar una fuerte correlación entre los valores de luminancia vecinos. Si nos fijamos por ejemplo en la figura de la esquina superior izquierda, observamos que si fijamos un valor para i_3 obtenemos la distribución de valores para i_4 conocido i_3 (lo cual corresponde a un corte vertical en el valor de i_3). Esta distribución depende del valor de i_3. Además, por la forma de esta distribución es muy probable que el valor de i_4 coincida con el de i_3 En el dominio ICA global, la correlación de segundo orden se ha eliminado, el valor esperado de c_m dado c_n es siempre cero independientemente del valor de c_n. Sin embargo, la varianza (o energía) del coeficiente c_m dado c_n depende del valor

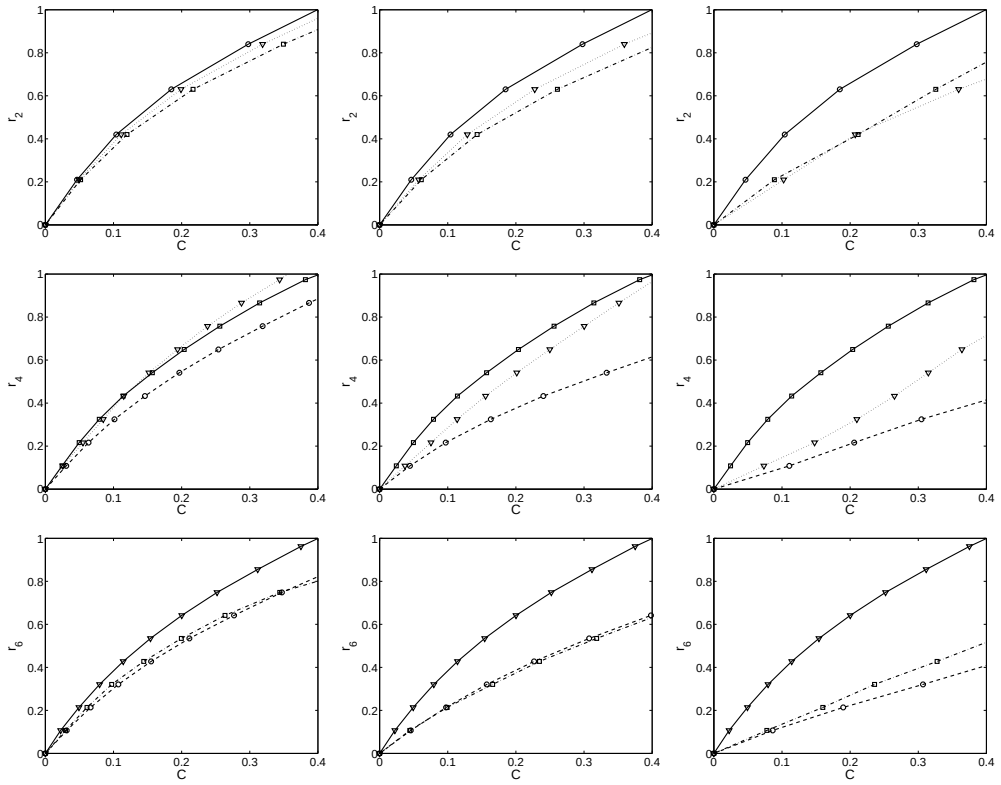


Fig. 3.11: Respuesta de los sensores de baja frecuencia (primera fila), frecuencia media (segunda fila) y alta frecuencia (tercera fila), cuando se utilizan diferentes estímulos de enmascaramiento de diferente contraste: contraste bajo (primera columna), contraste medio (segunda columna) y alto contraste (tercera columna). La respuesta sin enmascaramiento (línea continua) se muestra como referencia. Las demás curvas representan la respuesta del sensor considerado en presencia de un estímulo de enmascaramiento de baja (linea discontinua y círculos), media (línea punteada discontinua y cuadrados) y alta (línea punteada y triángulos) frecuencia.

concreto que tome c_n [24, 52]. En el dominio propuesto estas dependencias se eliminan casi por completo, mostrando de forma cualitativa que en este dominio existen menos dependencias que en el dominio ICA global lineal.

Estos resultados cualitativos son consistentes con los resultados cuantitativos de medidas de información mutua. La información mutua de una serie de variables $y_1, \cdots, y_n$ que toman valores discretos se define como la distancia de Kullback-Leibler entre la PDF conjunta de las variables y el producto de sus PDFs marginales, y puede obtenerse a partir de las entropías marginales, $H(y_i)$,

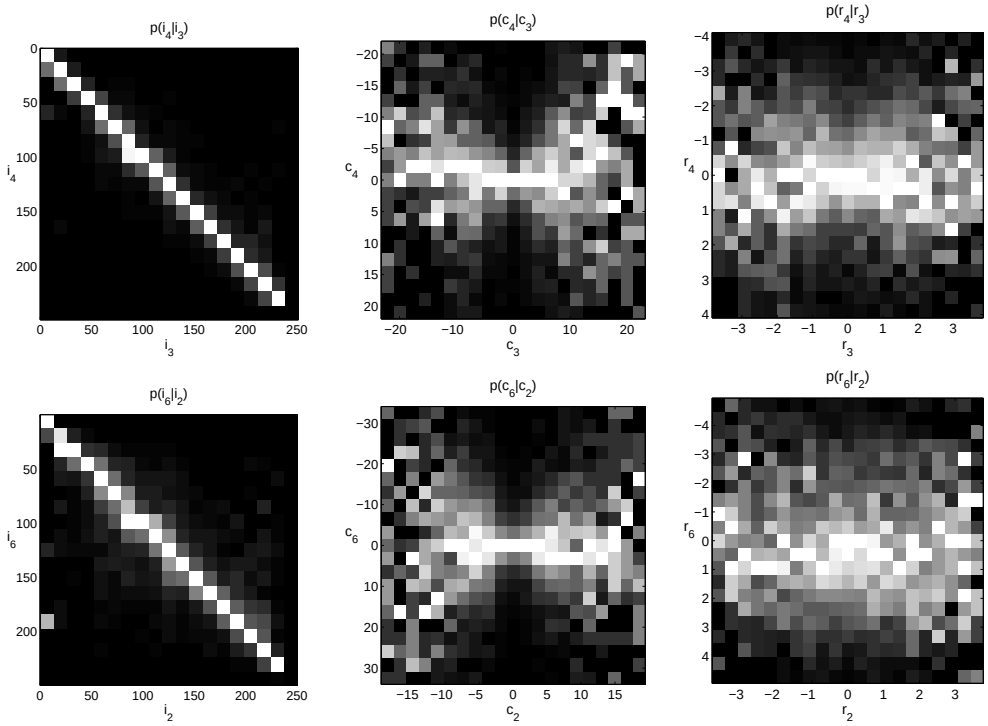


Fig. 3.12: Ejemplos de probabilidades condicionales de dos pares de coeficientes en los tres dominios considerados: dominio espacial (primera columna), dominio ICA global (segunda columna) y dominio propuesto (tercera columna).

y de la entropía conjunta, $H(y_1, \cdots, y_n)$, de las variables [61]:

$$I(y_1, \cdots, y_n) = \sum_{i=1}^{n} H(y_i) - H(y_1, \cdots, y_n). \tag{3.11}$$

La información mutua evalúa la información que proporcionan los vecinos sobre un coeficiente particular, y por tanto, lo fuertes que son las relaciones en un dominio particular.

Sin embargo, la información mutua depende de la entropía marginal de las variables consideradas. Un uso directo de la información mutua I para valorar las dependencias estadísticas en diferentes dominios (las entropías marginales serán diferentes en cada uno de los dominios, **i**, **c** y **r**) hará que el menor valor se produzca en el dominio en el que las entropías marginales sea menor. Para evitar esta tendencia, se define una medida de información mutua relativa que tiene en cuenta las entropías marginales [53]

$$I_r(y_1, \cdots, y_n) = \frac{\frac{1}{(n-1)} I(y_1, \cdots, y_n)}{\frac{1}{n} \sum_{i=1}^{n} H(y_i)}. \tag{3.12}$$

En este caso, $I_r = 1$ cuando los y_j son completamente redundantes y $I_r = 0$ cuando son completamente independientes.

La figura 3.13 muestra la información mutua relativa en los diferentes dominios en función de la distancia de las componentes (en dominio espacial mediante una línea punteada discontinua, en el dominio ICA global mediante una línea discontinua y en el dominio propuesto mediante una línea contínua).

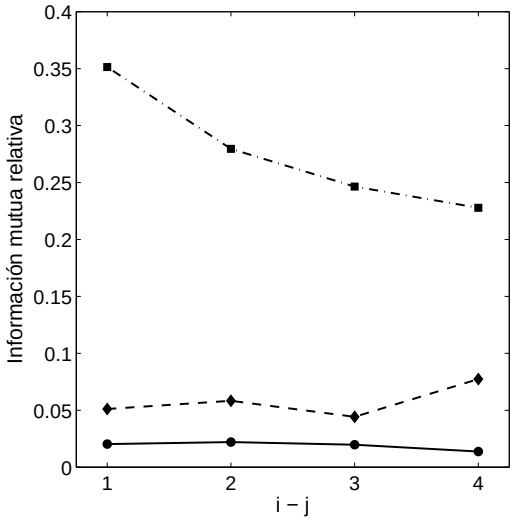


Fig. 3.13: Información mutua en los diferentes dominios, dominio espacial (línea punteada discontinua), en el dominio ICA global (línea discontinua) y en el dominio propuesto (línea continua), en función de la distancia entre las componentes

Como se puede observar, tanto en el análisis cualitativo como con el análisis cuantitativo, mediante el procedimiento descrito para la obtención de la representación en el dominio r se consigue que las componentes sean más independientes que en el dominio ICA global.

De estas tres subsecciones se deduce que la transformación obtenida buscando la independencia de las componentes de la representación (siguiendo la hipótesis de Barlow) da lugar a la reproducción de propiedades de las no linealidades de la respuesta de neuronas en V1 (tales como la saturación y el enmascaramiento cruzado).

3.5.6. Independencia del camino de integración

El ICA local propuesto asocia a cada punto, $\mathbf{i_0}$, un conjunto de vectores, $\mathbf{W}(\mathbf{i_0})$. Por lo tanto, la ecuación (3.8) puede interpretarse como la integración de un campo vectorial. El teorema de Stokes establece que la integración de un campo vectorial es independiente del camino de integración si la integral en circuitos cerrados es cero [62].

Para comprobar la independencia del camino de integración se han definido diferentes rutas cerradas (para diferentes contrastes) en diferentes planos como las ilustradas en la figura 3.14.

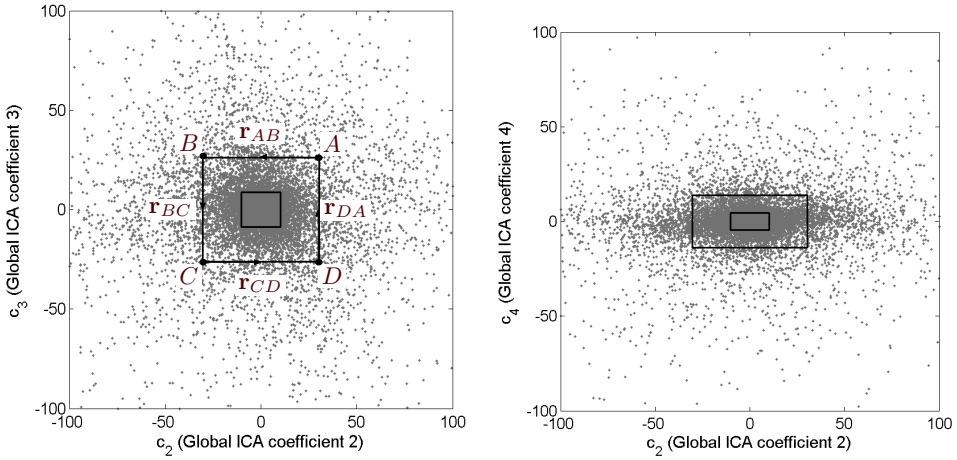


Fig. 3.14: Ejemplos ilustrativos de las rutas de integración cerradas utilizadas para comprobar la independencia del camino de integración.

Para estas rutas se ha calculado la respuesta de cada uno de los tramos que la definen, $\mathbf{r}_{AB}$, $\mathbf{r}_{BC}$, $\mathbf{r}_{CD}$, $\mathbf{r}_{DA}$, siendo la respuesta en el circuito cerrado la suma

$$\mathbf{r}_{ABCD} = \mathbf{r}_{AB} + \mathbf{r}_{BC} + \mathbf{r}_{CD} + \mathbf{r}_{DA}$$

Los siguientes vectores muestran la respuesta de los diferentes tramos para el circuito definido en el plano (c_2, c_3) mostrado en la figura figura 3.14.a,

$$r_{AB} = \begin{bmatrix} -0{,}43 \\ -2{,}20 \\ -1{,}10 \\ 0{,}01 \\ -0{,}18 \\ -0{,}39 \end{bmatrix} \quad r_{CD} = \begin{bmatrix} 0{,}59 \\ 2{,}00 \\ 1{,}10 \\ -0{,}05 \\ 0{,}19 \\ 0{,}33 \end{bmatrix} \quad r_{BC} = \begin{bmatrix} -0{,}52 \\ 1{,}40 \\ -1{,}50 \\ 0{,}33 \\ 0{,}25 \\ -0{,}69 \end{bmatrix} \quad r_{DA} = \begin{bmatrix} 0{,}47 \\ -1{,}40 \\ 1{,}60 \\ -0{,}07 \\ -0{,}45 \\ 0{,}56 \end{bmatrix} \tag{3.13}$$

	(c_2, c_4)	(c_2, c_3)	(c_3, c_4)	(c_3, c_5)	(c_4, c_6)
Contraste bajo	0.04	0.05	0.04	0.03	0.10
Contraste alto	0.12	0.10	0.09	0.05	0.09

Tab. 3.1: Normas relativas de las respuestas en las rutas consideradas

Se pueden apreciar una serie de propiedades interesantes en estas respuestas: por un lado, cuando nos movemos a lo largo del segmento que une los puntos A y B (reducción de la componente c_2) la mayor actividad se obtiene en la segunda componente del vector r_{AB}, sin embargo, este no es el único sensor activo revelando interacciones con el resto de coeficientes; por otro lado, si nos movemos en el segmento que va en el sentido opuesto, el vector de respuesta es aproximadamente el simétrico, $r_{CD} \approx -r_{AB}$. Esta tendencia también se cumple para los otros segmentos de forma que la suma final se aproxima a cero.

Se debe comprobar si $\mathbf{r}_{ABCD} = \mathbf{0}$. Sin embargo, si resulta que no es cero (como en el caso mostrado en (3.13)) debemos comprobar lo próximo que está al vector cero. Para ello se compara la norma de la respuesta $\mathbf{r}_{ABCD}$ con la suma de la normas de los diferentes tramos:

$$n = \frac{|\mathbf{r}_{ABCD}|}{|\mathbf{r}_{AB}| + |\mathbf{r}_{BC}| + |\mathbf{r}_{CD}| + |\mathbf{r}_{DA}|}$$

La tabla 3.1 muestra los resultados de esta norma relativa en los diferentes planos y caminos de integración.

Estos resultados muestran que la norma de la respuesta en rutas cerradas es pequeña (entre el 4 % y el 10 % de la norma de la respuesta media). Este resultado sugiere que el resultado de la integral (3.8) se puede considerar casi independiente del camino de integración y por lo tanto la integral a lo largo de diferentes rutas que una los puntos $E\{\mathbf{I}\}$ y $\mathbf{i}$ debe dar resultados bastante similares.

4. RELEVANCIA DE LA ESTIMACIÓN DEL MOVIMIENTO BASADA EN MODELOS NO LINEALES DE PERCEPCIÓN

En este capítulo se presentará el estudio realizado sobre las características que presenta el flujo óptico cuando se introducen aspectos perceptuales en el contexto de la codificación de video. Se comenzará revisando la estructura básica de los codificadores de vídeo basados en la compensación del movimiento. El objetivo de estos codificadores es formar una predicción del siguiente fotograma en la secuencia a partir del fotograma actual y la información de movimiento. El codificador envía la señal del error (diferencia entre la predicción y la imagen) transformada y cuantizada junto con la información del movimiento. El volumen de estas dos señales es intrínsecamente menor que el volumen de la secuencia original (al eliminar redundancias espaciales y temporales). Puesto que la eliminación de redundancias espaciales incluye cuantizadores perceptuales, los modelos de percepción influirán en los algoritmos de estimación adaptativa del movimiento.

Las aportaciones que se presentan en este capítulo son dos. Primero se estudia la relevancia relativa entre el módulo de la estimación del movimiento y el cuantizador. Para ello se probarán diferentes combinaciones de ambos y se estudiará cual es el efecto sobre la calidad de la secuencia reconstruida y cual es el error perceptual. Esto tendría interés debido a que se habían propuesto diferentes algoritmos óptimos [37, 63] pero no se habían comprobado subjetivamente (sólo numéricamente). Se muestra que la mejora del cuantizador es más relevante que la estimación del movimiento. No obstante, como segundo resultado, mostramos que la introducción de un modelo perceptual adecuado induce flujos más robustos en el contexto de segmentación. Para comprobar este hecho, se utilizará un algoritmo de segmentación a partir del flujo [38, 39] utilizando como entrada los diferentes flujos obtenidos. Además se propondrá una medida para evaluar la calidad de estas segmentaciones.

4.1. Estimación del movimiento en el contexto de la codificación de vídeo

4.1.1. Codificación de vídeo

Las secuencias de vídeo contienen gran cantidad de redundancia estadística y subjetiva en cada uno de los fotogramas que forman la secuencia y entre foto-

gramas consecutivos. El objetivo de la codificación de vídeo es la reducción de la cantidad de información necesaria para almacenar y transmitir teniendo en cuenta las redundancias estadísticas y subjetivas.

En una señal de vídeo hay dos tipos de redundancia estadística, la espacial y la temporal. Por esto hay dos tipos de codificación una llamada *intraframe* que intenta reducir la redundancia espacial y la llamada *interframe* que intenta reducir la redundancia temporal.

Con el fin de reducir la redundancia temporal, se aplica una forma particular de codificación predictiva llamada compensación del movimiento. El objetivo es predecir el siguiente fotograma de la secuencia a partir del fotograma actual y la información del desplazamiento de los objetos que forman la escena. Este desplazamiento se obtiene utilizando algoritmos de estimación de movimiento. De esta forma la señal de vídeo original se divide en dos señales: por una parte el desplazamiento (DVF *displacement vector filed*) y por otra parte el error cometido al realizar la predicción (DFD *displaced frame diference*).

La información del movimiento se codifica mediante DPCM y la señal de error se transforma y se cuantiza. La figura 4.1 muestra el esquema básico del codificador y del decodificador de vídeo basado en la compensación del movimiento [64, 65, 66].

Para reducir el error cometido en la predicción es importante realizar una buena estimación del movimiento. Sin embargo, la cantidad de información dedicada a representar este desplazamiento debe ser pequeña, lo cual es contradictorio con lo anterior. El propósito de las técnicas de estimación del movimiento en el contexto de la codificación de vídeo debe ser por tanto minimizar la cantidad de información dedicada a estas dos señales (DVF y DFD).

4.1.2. Estimación del movimiento con bloques de tamaño fijo

Los estándares MPEG-1, MPEG-2 y H-261 [65, 67] utilizan el algoritmo de correspondencia de bloques (BMA *Block-Matching Algorithms*) con un tamaño de bloque fijo (de tamaño 8×8) para realizar la estimación del movimiento.

En los métodos basados en la correspondencia de bloques se realiza una partición de la escena en regiones y se busca cual es la correspondencia de estas regiones en el siguiente fotograma. Para ello se define una medida de disparidad que será mínima para un determinado desplazamiento del bloque [64, 68]. Esta idea es ilustrada en la figura 4.2. Hay que indicar que diversos estándares de codificación de video como (MPEG-1, MPEG-2 y H.261) se basan en estas técnicas de tamaño de bloque fijo a pesar de que el algoritmo de estimación del movimiento no está especificado explícitamente.

La idea básica es que a todos los puntos que están dentro del bloque se les

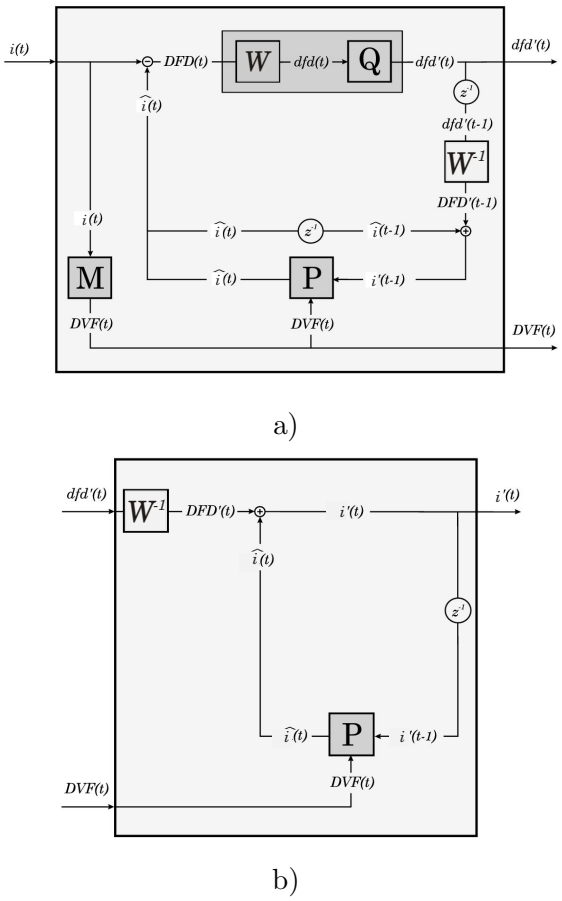


Fig. 4.1: Esquema general de un codificador, a), y un decodificador, b), de vídeo con compensación del movimiento. El codificador divide la secuencia de fotogramas de entrada, $\mathbf{i}(t)$, en una secuencia de información de movimiento, $DVF(t)$ y una secuencia de errores de predicción cuantizados en un dominio transformado, $dfd'(t)$. El esquema es un caso particular de cuantización predictiva, donde el objetivo es obtener una predicción del fotograma actual en función del fotograma anterior y la información de movimiento. Los símbolos z^{-1} representan retrasos en la señal en una unidad de tiempo discreto.

asigna el mismo desplazamiento. El modelo de movimiento implícito es el de objetos rígidos con movimiento traslacional, lo cual puede parecer una aproximación demasiado burda, pero hay que tener en cuenta que nada impide (a priori) dividir la imagen en bloques pequeños y que en cada uno de ellos se cumpla esta aproximación.

Para cada bloque, el vector de desplazamiento se evalúa realizando la correspondencia entre el bloque, que denotaremos como $\mathcal{B}$, en la imagen inicial y blo-

Fig. 4.2: El objetivo de un algoritmo de correspondencia de bloques es encontrar el desplazamiento que hay que aplicar a los bloques para encontrarlos en la escena un instante Δt posterior.

ques correspondiente en la imagen final dentro de una región de búsqueda, que denotaremos como $\mathcal{S}$, y buscando la posición espacial que minimiza el criterio

$$\mathbf{d} = \arg\min_{\mathbf{d}\in\mathcal{S}} \sum_{\mathbf{x}\in\mathcal{B}} \|i(\mathbf{x},t) - i(\mathbf{x}-\mathbf{d}, t+\Delta t)\| \tag{4.1}$$

Las medidas de distancia para medir esta disparidad son la norma cuadrática y el valor absoluto.

La figura 4.3 muestra un ejemplo de división en bloques de tamaño fijo (8×8) de la imagen de entrada.

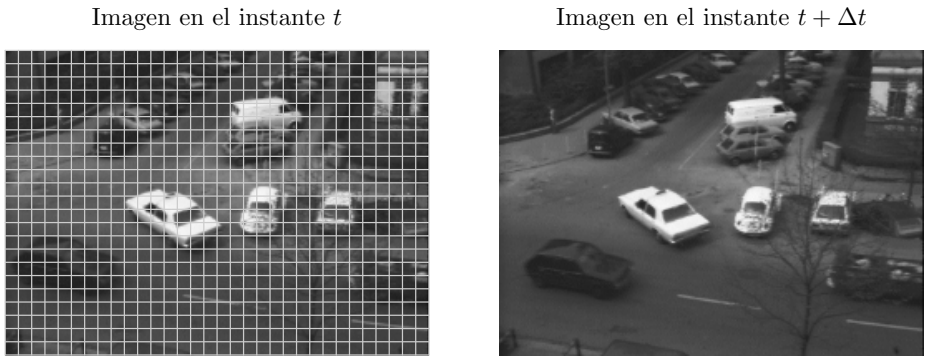


Fig. 4.3: Ejemplo de división de la imagen en bloques de tamaño 8×8

4.1.3. Estimación del movimiento adaptativa restringida por la entropía

El estándar H.263 introduce la posibilidad de utilizar un tamaño de bloque variable. El tamaño del bloque controla el balance entre lo fiable de la estimación del DVF que aumenta con el tamaño del bloque, y la capacidad de reducir la energía del DFD que disminuye con el tamaño del bloque. Por lo tanto, es conveniente poder ajustar el tamaño del bloque para adaptarse al contenido de la escena.

En este caso el proceso es iterativo: se comienza con la obtención del desplazamiento de regiones de tamaño grande, se dividen aquellas regiones en las que se cumpla un determinado criterio y se realiza la estimación del desplazamiento en cada una de las subdivisiones. La figura 4.4 es una ilustración del tipo de particiones que se obtienen con estos métodos.

Imagen en el instante t Imagen en el instante $t + \Delta t$

Fig. 4.4: Ejemplo de una posible división a la que llegaría un algoritmo de estimación jerárquica del movimiento con un tamaño de bloque inicial de 32×32 y un tamaño de bloque mínimo de 8×8

Como ya se ha comentado, en el contexto de la codificación de vídeo se transmiten dos señales: la señal de error de predicción transformada y cuantizada y los vectores de desplazamiento. Mediante esta descomposición se consigue representar la secuencia con un número menor de bits. Por tanto en este contexto tiene sentido proponer un criterio para la división de las regiones que tenga que ver con la cantidad de información que se obtiene con una determinada partición de la escena. Dufaux y Moscheni [36] propusieron el siguiente criterio: dividir un bloque en otros cuatro bloques para realizar una estimación más precisa del movimiento si el volumen de la información de las dos señales (DFD y DVF) es mayor antes de dividir.

$$H(\mathrm{DVF}_{\mathrm{split}}) + H(\mathrm{DFD}_{\mathrm{split}}) < H(\mathrm{DVF}_{\mathrm{no\ split}}) + H(\mathrm{DFD}_{\mathrm{no\ split}}) \tag{4.2}$$

donde $H(\text{DVF})$ es la entropía del DVF codificado con un DPCM más la información necesaria para codificar la estructura del *quadtree* y $H(DFD)$ es la entropía de la diferencia entre el fotograma actual y la predicción (sin transformar ni cuantizar).

4.1.4. Estimación del movimiento restringida por la entropía perceptual

Utilizando la medida de entropía perceptual, Malo *et al.* [37] propusieron un criterio de división de bloques en el que se dividen regiones sólo si la información del movimiento que se obtiene tras la división da lugar a una información de movimiento perceptualmente relevante.

El argumento esgrimido consistía en que un criterio de división de bloques basado en la entropía en el dominio espacial (como el propuesto por Dufaux *et al* [36]) puede que no mejore la calidad percibida de la secuencia reconstruida, ya que la reducción de la entropía de la señal de error en el dominio espacial no tiene en cuenta la cuantización (que es dependiente de la frecuencia). El esfuerzo de refinamiento en la estimación del movimiento se debería realizar en las regiones que contengan *información de movimiento perceptualmente significativa*.

El criterio que propusieron para dividir un bloque en la descomposición quadtree fue el siguiente

$$H(\text{DVF}_{\text{split}}) + H_p(\text{DFD}_{\text{split}}) < H(\text{DVF}_{\text{no split}}) + H_p(\text{DFD}_{\text{no split}}) \tag{4.3}$$

donde $H(\text{DVF})$ es la entropía del DVF codificado con un DPCM más la información necesaria para codificar la estructura del *quadtree* y H_p es la entropía perceptual definida en la ecuación 4.4.

Según este criterio, el esquema óptimo consiste en utilizar un codificador de vídeo que utilice un cuantizador Q_p basado en un modelo perceptual y que utilice la medida de la entropía H_p en el criterio de división de bloques.

La figura 4.5 muestra ejemplos de los diferentes flujos obtenidos utilizando los tres métodos comentados para la secuencia del cubo y del taxi.

La utilización de modelos no lineales en estimación de flujo se refiere a la forma de medir la entropía perceptual usando Q_p.

4.1.5. Entropía perceptual y cuantizadores

En los trabajos de Watson [5] y Daugman [69] sobre la reducción de la entropía en el sistema visual humano se asume un modelo de procesado del contraste

Fig. 4.5: Ejemplos de flujo con los diferentes métodos comentados. a) método de estimación del movimiento con tamaño de bloque uniforme, b) con tamaño de bloque variable utilizando en el criterio de división la entropía en el dominio espacial y c) con tamaño de bloque variable utilizando en el criterio de división la entropía perceptual.

basado en el par (W, Q_p), donde W es un banco de filtros y Q_p es un cuantizador perceptual, y sugieren que la entropía de la representación cortical (una medida de la entropía perceptual de la señal) es la entropía de la versión cuantizada de la imagen transformada:

$$H_p(\mathbf{i}) = H(Q_p(\mathbf{W} \cdot \mathbf{i})) \tag{4.4}$$

En el siguiente apartado revisaremos cómo se se pueden obtener cuantizadores basados en modelos de respuesta del sistema visual humano.

Cuantizadores basados un modelo de respuesta

La forma natural de evaluar la calidad de una imagen o secuencia codificada es mediante la comparación uno a uno entre la señal codificada y la señal original. El resultado de esta comparación entre la señal original y la codificada está relacionada con la capacidad del observador de detectar el ruido de cuantización.

Las técnicas convencionales para el diseño de cuantizadores para transformadas utilizan una medida de distorsión media de forma que el cuantizador final logra un error medio mínimo sobre el conjunto de datos de entrenamiento.

Sin embargo, la minimización del error medio no garantiza que para alguna señal particular se obtenga un buen resultado desde el punto de vista subjetivo.

Para evitar errores perceptualmente significativos en señales particulares, el codificador debe restringir el error perceptual máximo (MPE) en cada coeficiente y amplitud. Este requisito se cumple si se distribuyen los niveles de forma uniforme en el dominio r (lo cual resulta en una distribución de niveles no uniforme de los niveles en el dominio a si se utiliza un modelo de respuesta no lineal, véase la figura 2.2).

Este criterio ha sido utilizado para la compresión de imágenes para conseguir una contribución constante al error para cada componente frecuencial [41, 27].

El diseño de los cuantizadores para un conjunto de coeficientes de la transformada por bloques consta de dos pasos; en primer lugar hay que encontrar la distribución de los niveles de cuantización y en segundo lugar hay que encontrar el número de niveles por coeficiente (*bit allocation*).

Si asumimos un modelo de respuesta no lineal sin interacción entre los coeficientes (ecuación 2.4) la distancia entre dos patrones (en el dominio transformado) $\mathbf{c}$ y $\mathbf{c} + \Delta\mathbf{c}$ viene dada por:

$$D^2(\mathbf{c}, \mathbf{c} + \Delta\mathbf{c}) = \sum_f D_f^2 = \sum_f G_f(c_f)\Delta c_f^2$$

donde $G_f(c_f)$ viene dada por la expresión 2.5.

Con el fin de evitar grandes errores perceptuales en bloques individuales el codificador debe ser diseñado de forma que se acote el error perceptual máximo para cada frecuencia f y amplitud c_f.

Si a un coeficiente (con frecuencia f) se le van a asignar N_f niveles de cuantización distribuidos según una densidad $\lambda_f(c_f)$, el error Euclídeo máximo en la amplitud a_f estará acotado por la mitad de la distancia Euclídea entre dos niveles de cuantización

$$\Delta c_f(c_f) =\leq \frac{1}{2N_f\lambda_f(c_f)}$$

Sustituyendo esta ecuación en la anterior se obtiene que el error perceptual máximo para esta frecuencia y amplitud está relacionado con la métrica y con la densidad de niveles

$$\mathrm{MPE}_f(c_f) = G_f(c_f)\,\mathrm{máx}(\Delta c_f(c_f))^2 = \frac{G_f(c_f)}{4N_f^2\lambda_f^2(c_f)}$$

Por lo tanto, la densidad de niveles de cuantización que da lugar a una cota constante del error perceptual máximo en todo el rango de amplitud es

$$\lambda_{f_{\mathrm{opt}}}(c_f) = \frac{G_f(c_f)^{\frac{1}{2}}}{\int G_f(c_f)^{\frac{1}{2}}\mathrm{d}c_f} \tag{4.5}$$

Tal y como vimos en el capítulo 2, un determinado modelo de respuesta induce una métrica que viene dada por la ecuación $G(\mathbf{c}) = \nabla R(\mathbf{c})^T\nabla R(\mathbf{c})$, y por tanto, $W(\mathbf{c})^{\frac{1}{2}} = \nabla R(\mathbf{c})$, es decir, la densidad de niveles óptima viene dada por la sensibilidad del sistema (se deben asignar más niveles de cuantización donde el sistema es más sensible).

Con estas densidades óptimas, el error perceptual máximo para cada coeficiente f dependerá del número de niveles asignado y del valor integrado de la raíz cuadrada de la métrica,

$$\mathrm{MPE}_{f_{\mathrm{opt}}} = \frac{1}{4N_f^2}\Big(\int G_f(c_f)^{\frac{1}{2}}\mathrm{d}c_f\Big)^2$$

Fijando la misma distorsión para cada coeficiente $\mathrm{MPE}_{f_{\mathrm{opt}}} = k^2$, podemos obtener el número óptimo de niveles de cuantización, N_f como

$$\mathrm{N}_{f_{\mathrm{opt}}} = \frac{1}{2k} \int G_f(c_f)^{\frac{1}{2}} \mathrm{d}c_f \tag{4.6}$$

Una vez obtenido el número de niveles $N_{f_{\mathrm{opt}}}$ utilizando la ecuación 4.6 habría que distribuir los niveles según la densidad $\lambda_{f_{\mathrm{opt}}}(c_f)$ utilizando la ecuación 4.5. Tanto $N_{f_{\mathrm{opt}}}$ como $\lambda_{f_{\mathrm{opt}}}(c_f)$ dependen de la métrica. Como ya se ha comentado la métrica utilizada es la inducida por el modelo no lineal de respuesta sin interacción entre coeficientes (véase la figura 2.3).

La figura 4.6 muestra un ejemplo de reparto de niveles por coeficiente y de cuantizadores para dos coeficientes de diferente frecuencia para dos valores del error.

Si se hubiese tomado un modelo de respuesta lineal (ecuación 2.3) se habrían obtenido cuantizadores basados en la CSF del estilo utilizado en MPEG. En este caso se obtendrían cuantizadores uniformes para cada coeficiente (véase la figura 2.1) y N_f sería proporcional a la CSF, que es una de las opciones recomendadas en JPEG y MPEG.

En las dos secciones siguientes se comprueba si es relevante y en qué aspectos el flujo obtenido utilizando el criterio de división de bloques basado en la entropía perceptual en el contexto de la codificación de video.

4.2. Relevancia relativa flujo-cuantización

En esta sección se va a estudiar si la calidad de la secuencia reconstruida mejora cuando se utiliza el criterio de división por bloques basado en la entropía perceptual. Para ello se han desarrollado una serie de experimentos que se describen a continuación.

4.2.1. Experimentos para comprobar la relevancia relativa de las mejoras en la estimación del movimiento y en la cuantización

Para estudiar las ventajas relativas de las mejoras propuestas (en el cuantizador y en la estimación del movimiento) las cuatro combinaciones posibles de algoritmos de cuantización y estimación del movimiento se comparan a una misma tasa de compresión.

Las combinaciones son las que se detallan a continuación:

- *Perceptually weighted VSBMA* con un cuantizador MPE basado en la respuesta lineal (tipo MPEG).

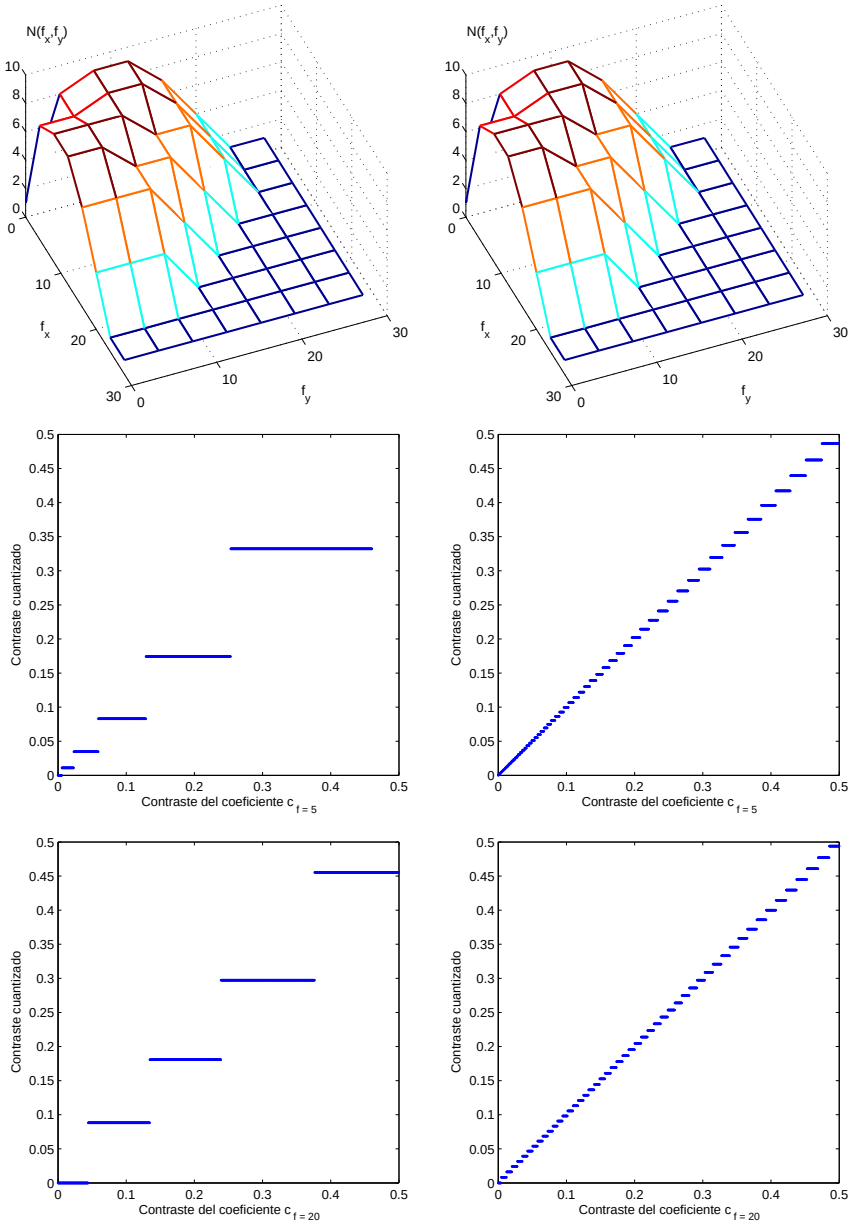


Fig. 4.6: En cada columna se muestra la distribución de niveles por frecuencia, el cuantizador para una frecuencia $f = 5$ cpg y el cuantizador para una frecuencia $f = 20$ cpg. a) Para un error $k = 0{,}1$ y b) para un error $k = 0{,}01$.

- *Perceptually weighted VSBMA* con un cuantizador MPE basado en la respuesta no lineal sin interacción entre coeficientes.

- *Unweighted VSBMA* con un cuantizador MPE basado en la respuesta lineal (tipo MPEG).
- *Unweighted VSBMA* con un cuantizador MPE basado en la respuesta no lineal sin interacción entre coeficientes.

El primer análisis consiste en comprobar qué porcentaje del total (DFD y DVF) se dedica al DVF al utilizar cada uno de los métodos de estimación del flujo. La tabla 4.1 muestra los resultados para cuatro secuencias. Como se puede observar el método propuesto por Malo *et al* [37] proporciona un flujo que se puede codificar con menos bits que el resto.

Tab. 4.1: Volumen relativo del movimiento (en %) respecto al volumen total de la señal codificada (200 kb/s)

	TAXI	RUBIK	YOSEMITE	TREES	Media
FSBMA	12.57	21.71	46.19	39.21	30 ± 8
Suboptimal VSBMA	1.89	3.29	8.69	4.97	4.7 ± 1.5
Optimal VSBMA	0.94	1.87	4.32	3.24	2.6 ± 0.7

El segundo análisis consiste en comprobar la calidad de la secuencia decodificada mediante inspección visual y mediante el error cuadrático medio (MSE) y el error cuadrático medio perceptual (PMSE), utilizando la métrica dada en la ecuación 2.5. La figura 4.7 muestra los errores de las cuatro combinaciones estudiadas en cada fotograma. La figura 4.8 muestra el fotograma número 7 en la secuencia decodificada.

Estos resultados confirman las ventajas cuantitativas del esquema óptimo: el volumen del flujo y la distorsión (independientemente de la medida) se reducen. Sin embargo esta disminución es casi despreciable en relación con las ventajas de una mejor cuantización. Esto se puede observar en las curvas de la figura 4.7 y lo que es más importante, en la calidad de las secuencias reconstruidas 4.8.

Sin embargo, tal y como mostraremos en la siguiente sección, la información sobre el movimiento obtenida mediante el esquema propuesto tiene además otras ventajas.

4.3. *Flujo perceptual para segmentación*

El flujo óptico sirve como entrada a los métodos de segmentación de objetos basados en el flujo óptico que producen como resultado una división de la escena

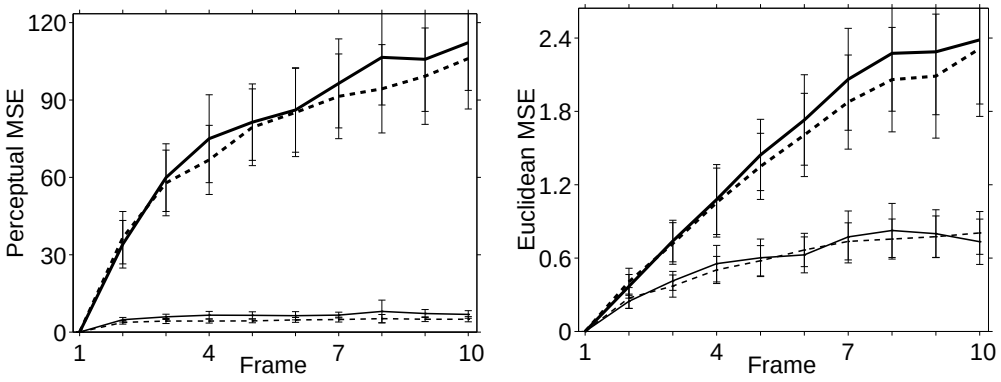


Fig. 4.7: Error (MSE y Perceptual MSE) cometidos por los diferentes esquemas de codificación para la secuencia RUBIK. Linea continua gruesa cuantizador MPEG y flujo adaptativo subóptimo. Linea discontinua gruesa MPEG y flujo adaptativo óptimo. Línea continua delgada cuantizador no lineal y flujo adaptativo subóptimo. Línea discontinua delgada cuantizador no lineal y flujo adaptativo óptimo.

en objetos, agrupando regiones con movimiento semejante. Una posible forma de evaluar un método de estimación del flujo óptico podría ser observando la calidad de la segmentación que induce. En esta sección se muestran las segmentaciones obtenidas con los diferentes flujos.

4.3.1. Algoritmo VOPs

En concreto se ha utilizado una representación en capas [38, 39]. Para obtener esta representación se parte de los vectores de desplazamiento proporcionados por un método de estimación del movimiento determinado.

La aproximación descrita en [39] se basa en comenzar con una partición en regiones. En cada región se ajusta un modelo afín [70] que proporciona 6 parámetros para cada región, $\mathbf{a}$, estos parámetros definen la velocidad de cada punto en la región. A continuación se agrupan las regiones cuya distancia entre modelos afines sea mínima. El proceso se itera mientras la distancia entre modelos esté por debajo de un determinado umbral. La medida de distancia entre modelos se define como

$$\mathrm{dist}(\mathbf{a}_i, \mathbf{a}_j) = \frac{1}{n} \sum_{\mathbf{x} \in \mathcal{R}_i, \mathcal{R}_j} \|V_{a_i}(\mathbf{x}) - V_{a_j}(\mathbf{x})\|^2$$

es decir, dos modelos afines tienen distancia pequeña si los vectores de desplazamiento que generan son parecidos.

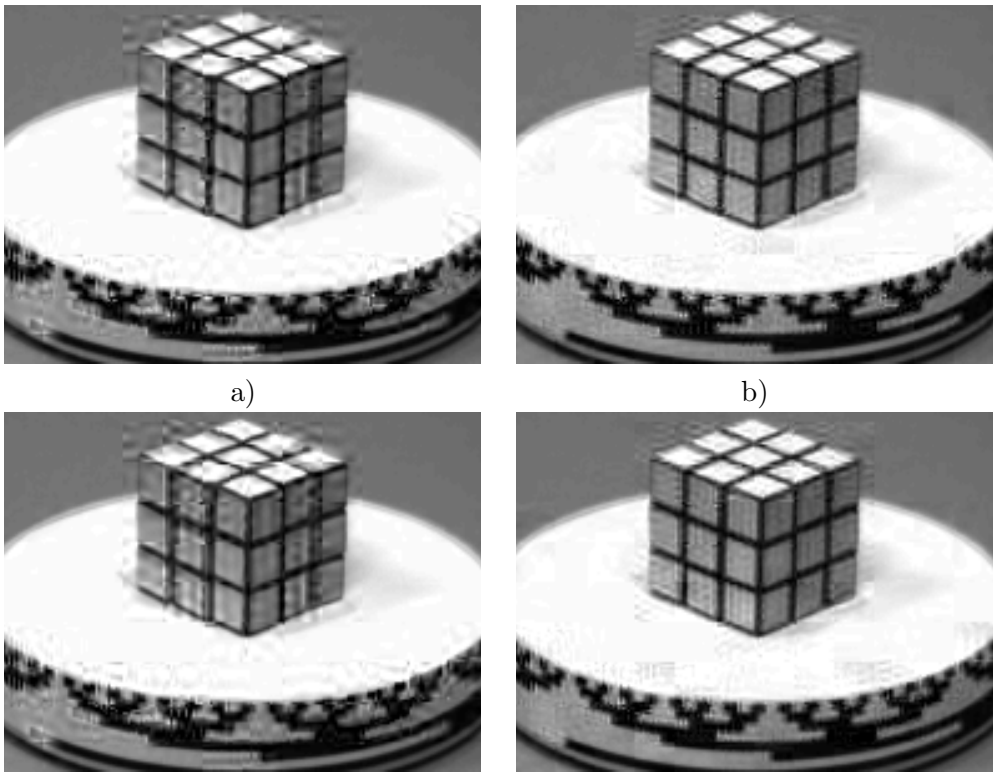


Fig. 4.8: Fotograma 7 de la secuencia RUBIK reconstruida para los diferentes esquemas de codificación. a), b), c) y d)

Dado un par de imágenes consecutivas de la secuencia con n objetos en movimiento, el algoritmo de segmentación proporciona m máscaras representado los objetos identificados y los modelos afines correspondientes.

4.3.2. Medida de la calidad de la segmentación

Con el fin de proporcionar una evaluación objetiva de la calidad de la segmentación se propuso una medida basada en la distancia *chamfer* [40]. La distancia *chamfer* mide la distancia que hay desde cualquier punto de una imagen al borde más cercano. La medida del error, ϵ_i, al representar la segmentación (realizada de forma manual) del objeto O_i mediante una máscara M (M es la máscara que maximiza $M_j \cap O_i$ para todas las máscaras obtenidas M_j) se define como

$$\epsilon_i = \int_{M \cap O_i^C} D_i(x)\, dx + \int_{O_i \cap M^C} D_i(x)\, dx \tag{4.7}$$

Tab. 4.2: Errores de segmentación, número de iteraciones y número de regiones identificadas.

	RUBIK (2 objects)			TAXI (4 objects)		
	ϵ	Iterac.	Regiones	ϵ	Iterac.	Regiones
FSBMA	1.00 ± 0.05	25	21 ± 5	1.00 ± 0.14	25	15 ± 4
Unweighted VSBMA	0.46 ± 0.06	21 ± 5	4 ± 2	0.52 ± 0.15	11 ± 9	6 ± 2
Perceptual VSBMA	0.27 ± 0.03	7 ± 6	2.1 ± 0.3	0.49 ± 0.14	4 ± 1	4.3 ± 0.5

donde C denota conjunto complementario, $D_i(x)$ denota la distancia *chamfer*. El error cometido por la segmentación, ϵ, es la suma de los errores de segmentación de cada objeto.

4.3.3. Experimentos y resultados

Utilizando el procedimiento descrito en la sección 4.3.1, podemos comprobar que la segmentación resultante es mejor cuando se utiliza la información de movimiento obtenida mediante el esquema propuesto.

Las medidas del error cometido (obtenido utilizando la ecuación 4.7 cuando se segmenta la escena utilizando los diferentes flujos se muestran en la tabla 4.2.

Las segmentaciones obtenidas con los diferentes flujos se muestran en la figura 4.9. Como puede verse en la figura 4.9 y en la tabla 4.2 el algoritmo perceptual no sólo minimiza el error de segmentación sino que además acelera la convergencia del algoritmo de segmentación porque da lugar a un flujo más significativo y coherente. Esta segmentación se podría refinar utilizando el perímetro de cada máscara como el contorno inicial de un *snake* [71, 72, 73, 74].

Fig. 4.9: Ejemplos de segmentaciones a partir de los diferentes flujos. a) Fotograma original, b) Segmentación manual y segmentaciones obtenidas utilizando un método de estimación del movimiento c) con tamaño de bloque uniforme, d) con tamaño de bloque variable utilizando en el criterio de división la entropía en el dominio espacial y e) con tamaño de bloque variable utilizando en el criterio de división la entropía perceptual.

5. RESTAURACIÓN BASADA EN MODELOS NO-LINEALES DE PERCEPCIÓN

En este capítulo se revisará la teoría de la regularización de problemas inversos mal planteados y se mostrará la formulación de la restauración de imágenes degradadas como un caso particular de regularización en un dominio discreto. Se revisará el método de la L-curva para la obtención del parámetro de regularización y se propondrá una modificación. Se propondrán operadores de penalización basados en la respuesta no lineal del sistema visual humano. Esta clase de operadores se comparará con otros operadores clásicos no adaptativos (como por ejemplo el operador basado en la segunda derivada y el operador basado en la CSF) y otros adaptativos (basados en modelos autorregresivos de diferente orden). Finalmente se realizará un estudio del comportamiento de los diferentes operadores proponiendo una medida de sensibilidad de la solución frente a variaciones del parámetro de regularización al utilizar un determinado operador, también se propondrá una medida de robustez de un operador frente a la degradación y finalmente se comprobará su validez en la tarea de restauración de imágenes degradadas.

5.1. Problemas mal planteados

El problema de determinar una solución a la ecuación $f = H(g)$ en un espacio métrico F con métrica $\rho_F(\cdot,\cdot)$ a partir de datos disponibles g en un espacio métrico G con métrica $\rho_G(\cdot,\cdot)$, se dice que está bien planteado si se cumplen las tres condiciones siguientes:

1. Para todo $g \in G$ existe una solución $f \in F$.
2. La solución es única.
3. El problema es estable en los espacios F y G, es decir, para todo $\epsilon > 0$ existe un $\delta(\epsilon) > 0$ tal que $\forall\, g_1, g_2 \in G$ se cumple que si $\rho_G(g_1, g_2) \leq \delta(\epsilon)$ entonces $\rho_F(f_1, f_2) \leq \epsilon$ con $f_1 = H(g_1)$ y $f_2 = H(g_2)$. En otras palabras, la transformación inversa H^{-1} debe ser continua. Con esto se exige que pequeñas perturbaciones en los datos se transformen en pequeñas perturbaciones en la solución.

Si el problema no cumple alguna de estas condiciones se dice que está **mal planteado**. El concepto de problema bien planteado fue introducido por Jacques Hadamard en un trabajo publicado en 1902 [75] y revisado posteriormente en 1923 [76], quien adoptó el punto de vista de que todo problema matemático correspondiente a problemas reales debe ser bien planteado, en caso contrario es un problema que no tiene sentido. Esto redujo el interés sobre los problemas inversos mal planteados.

Sin embargo, una amplia clase de los llamados problemas inversos que aparecen en física, matemáticas y otras ramas de la ciencia, pertenecen a la clase de problemas mal planteados en el sentido de Hadamard. Supongamos que en el modelo matemático para un determinado experimento, el objeto a estudiar (la variable a determinar f, que puede ser una función o un vector) pertenece a un conjunto F de posibles soluciones. Supongamos que f no puede ser medido directamente y que lo que se puede medir es

$$g_o = Hf_o$$

donde el subíndice o indica que se trata de las señales originales. Supongamos ahora que debido al proceso de medición se obtiene g_d en lugar de g_o, donde el subíndice d indica que se trata de una señal que ha sufrido alguna degradación. El objetivo es obtener una estimación de f_o a partir de la observación g_d. Bajo estas condiciones encontramos que hay que obtener una solución para f en la ecuación

$$Hf = g_d \tag{5.1}$$

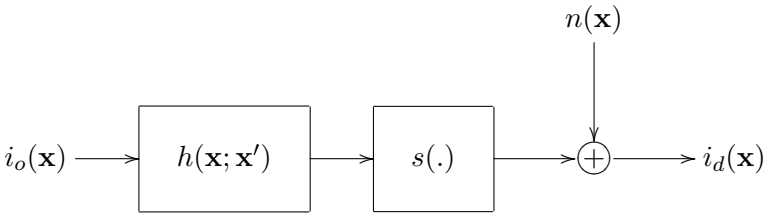
que será una aproximación a f_o. Entonces puede ocurrir lo siguiente:

- La solución puede no existir en F ya que puede ser que g_d no esté en el rango del operador H
- La solución puede no ser única si el núcleo de H^{-1}, $N(H^{-1})$, es no vacío. En ese caso, si f es una solución y $g_N \in N(H^{-1})$ entonces también será una solución $f + H^{-1}g_N$.
- La solución puede que no sea estable bajo pequeñas perturbaciones de g_d debido al hecho de que H^{-1} no sea continua.

Por tanto el problema 5.1 estaría mal planteado. La mayoría de problemas inversos son mal planteados [77, 78].

5.2. La restauración como problema mal planteado

Como ejemplo supongamos que una escena pasa a través de un sistema óptico y es almacenada. Si el proceso de formación de la imagen es lineal, la imagen grabada puede ser modelada como la salida del sistema mostrado en el siguiente diagrama [28, 79, 80]



donde $i_o(\mathbf{x})$ representa la imagen original, $i_d(\mathbf{x})$ denota la imagen resultante, $h(\mathbf{x};\mathbf{x}')$ es la respuesta impulsional 2-D del sistema lineal de degradación y $s(\cdot)$ representa la no-linealidad del sensor. La contribución del ruido $n(\mathbf{x})$ se suele modelar como un proceso aleatorio aditivo que no está correlacionado con la imagen.

Si se asume que la respuesta impulsional es estacionaria, que se está en un dominio discreto y se asume que el sensor es lineal, entonces el modelo simplificado resultante es [28]

$$i_d(\mathbf{m}) = h(\mathbf{m}) \star i_o(\mathbf{m}) + n(\mathbf{m}) \tag{5.2}$$

donde $\star$ indica convolución y $\mathbf{m} = (m_1, m_2)$ son los índices de la matriz. Esta ecuación se puede poner en forma matricial

$$\mathbf{i}_d = \mathbf{H}\mathbf{i}_o + \mathbf{n} \tag{5.3}$$

donde $\mathbf{i}_o$, $\mathbf{i}_d$ y $\mathbf{n}$ son los vectores resultantes de la ordenación lexicográfica de las matrices correspondientes y $\mathbf{H}$ es el operador de emborronamiento.

La ecuación 5.2 se puede expresar en el dominio de Fourier,

$$c_d(\boldsymbol{\nu}_\mathbf{m}) = H(\boldsymbol{\nu}_\mathbf{m})c_o(\boldsymbol{\nu}_\mathbf{m}) + \mathcal{N}(\boldsymbol{\nu}_\mathbf{m}) \tag{5.4}$$

donde c_d, c_o y $\mathcal{N}$ representan a las transformadas de Fourier de la imagen degradada, de la imagen original y del ruido respectivamente y $\boldsymbol{\nu}_\mathbf{m} = (\nu_{m1}, \nu_{m2})$ denota frecuencia espacial.

Supongamos ahora que no existe ruido y que la degradación consiste en un filtro pasa-baja de una determinada frecuencia de corte. Una vez observada $i_d(\mathbf{m})$ la solución trivial consiste en obtener una estimación, $i'(\mathbf{m})$, de $i_o(\mathbf{m})$ como

$$i'(\mathbf{m}) = \mathcal{F}^{-1}\left(\frac{c_d(\boldsymbol{\nu}_\mathbf{m})}{H(\boldsymbol{\nu}_\mathbf{m})}\right) \tag{5.5}$$

donde $\mathcal{F}^{-1}$ denota la transformada inversa de Fourier.

Para poner un ejemplo concreto supongamos que el filtro pasa-baja tiene una frecuencia de corte de $f_c = 4$ cpd (ciclos por grado)[1]. La imagen degradada observada se muestra en la figura 5.1.a. Aplicando la ecuación 5.5 se obtiene como resultado la imagen que se muestra en la figura 5.1.b.

Supongamos ahora que la degradación, además del emborronamiento incluye ruido blanco aditivo de varianza $\sigma^2 = 5$. La imagen degradada observada se muestra en la figura 5.1.c.

En este caso aplicando la ecuación 5.5 se obtiene como resultado la imagen que se muestra en la figura 5.1.d, es decir, existe una solución al sistema de ecuaciones pero esta solución no es aceptable. Por lo tanto, podemos concluir que el problema de la restauración es un problema mal planteado ya que una pequeña variación en los datos observados provoca una gran variación en la solución.

Este resultado es fácilmente explicable realizando un análisis mediante la descomposición en valores singulares de la matriz de emborronamiento. Supongamos una imagen de tamaño 16×16, y un emborronamiento con un filtro pasa-baja con frecuencia de corte $f_c = 4$ cpd, cuya respuesta impulsional se muestra en la figura 5.2.a. Expresando esta respuesta impulsional como un operador que actúa sobre un espacio de dimensión 256×1 y produce un resultado en un espacio de dimensión 256×1, se obtiene una matriz cuya estructura se muestra en la figura 5.2.b.

La descomposición en valores singulares de un operador da como resultado tres matrices de tal forma que,

$$H = USV^T \tag{5.6}$$

donde se cumple que $UU^T = I$, $VV^T = I$ y S es una matriz diagonal. De estas propiedades se deduce que

$$H^{-1} = VS^{-1}U^T \tag{5.7}$$

[1] Estamos asumiendo una frecuencia de muestreo de 64 cpd, es decir una frecuencia de Nyquist de 32 cpd.

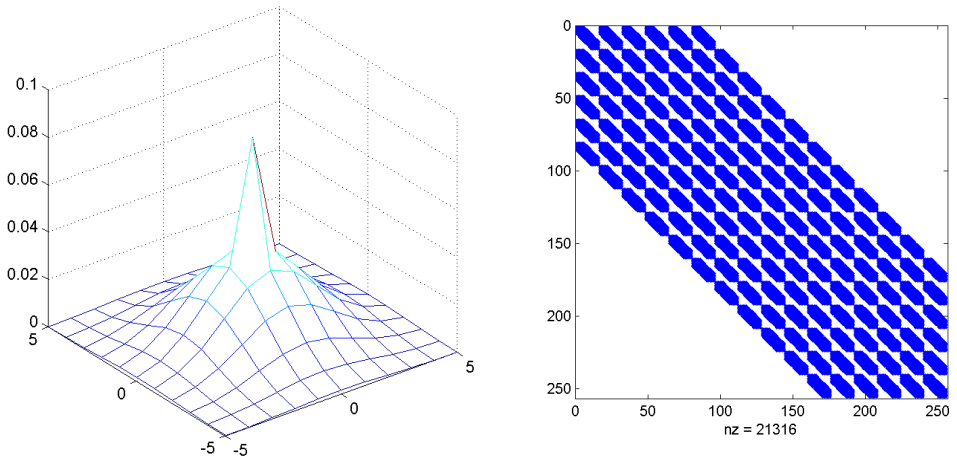
a) b)

c) d)

Fig. 5.1: Ejemplo de imágenes degradadas (con y sin ruido) y restauradas realizando la inversión del operador de emborronamiento. a) Imagen degradada observada cuando la degradación consiste únicamente en emborronamiento con un filtro pasa baja con frecuencia de corte 4 cpg. (PSNR=23.9092 db). b) Imagen restaurada aplicando la ecuación 5.5 (PSNR=43.4413 db). c) Imagen degradada observada cuando la degradación consiste en emborronamiento con un filtro pasa baja con frecuencia de corte de 4 cpg y ruido aditivo de varianza 5 (PSNR=23.8258 db). d) Imagen restaurada aplicando la ecuación 5.5 (PSNR=15.3279 db).

Para el ejemplo concreto, las matrices U, S y V^T se muestran en la figura 5.3.

La matriz V^T se puede interpretar como un análisis de la señal (producto escalar de la señal por una serie de funciones base), a los coeficientes obtenidos



a) b)

Fig. 5.2: a) Respuesta impulsional, h, de un filtro pasa baja con frecuencia de corte $f_c = 4$ cpd. b) Estructura de un operador equivalente, H (con ordenación lexicográfica)

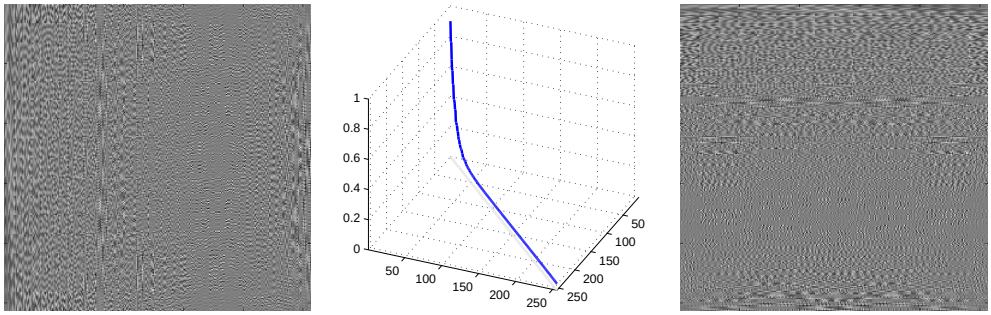


Fig. 5.3: Matrices U, S y V^T de la descomposición en valores singulares de la matriz H.

se les aplica una ganancia que viene dada por la diagonal de la matriz S y finalmente una etapa de síntesis realizada por la matriz U que nos lleva de vuelta al dominio original. ¿Cuales son estas funciones base? Si mostramos las filas de V^T reordenadas como imágenes 16×16 se obtienen las funciones mostradas en la figura 5.4

Puesto que los valores singulares mayores están en las primeras posiciones, resulta que de la descomposición de la señal en estas funciones base nos estamos quedando con las de baja frecuencia, es decir se están atenuando las de alta frecuencia (como cabía esperar por tratarse de un filtro pasa baja).

Cuando se quiere obtener i' y se aplica H^{-1} la situación cambia y las funciones

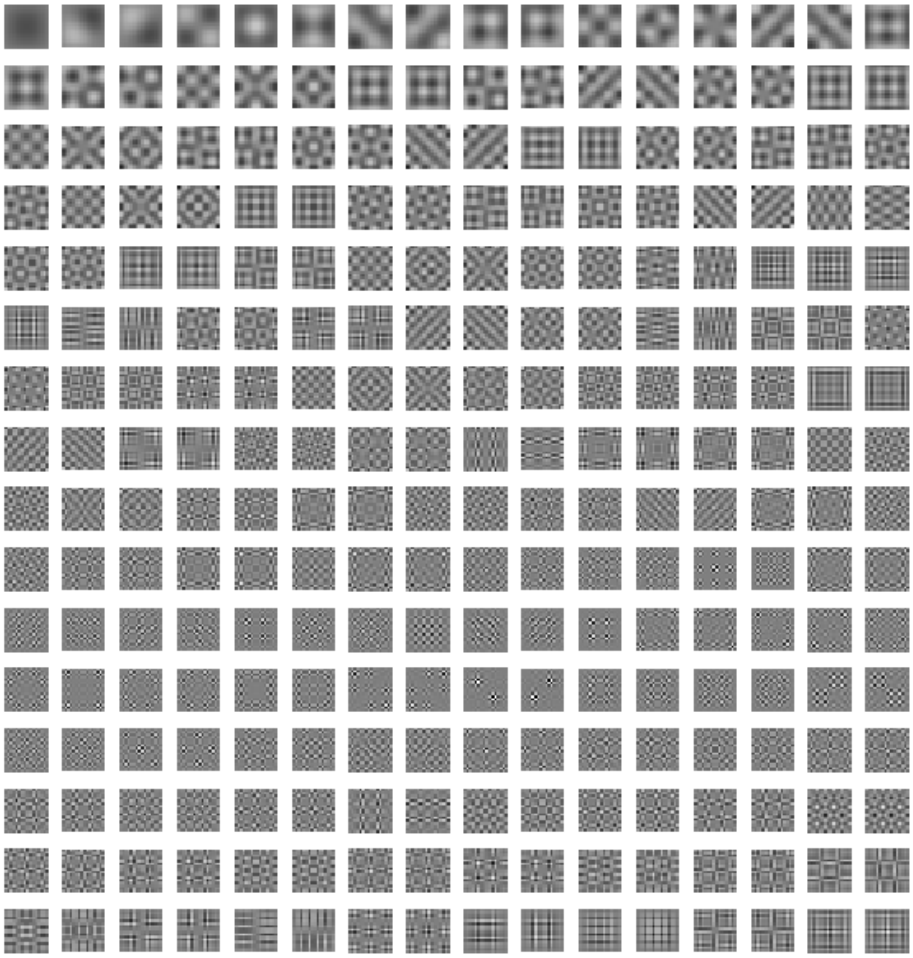


Fig. 5.4: Funciones base de la etapa de análisis de la descomposición en valores singulares (filas de la matriz V^T reordenadas para obtener imágenes de tamaño 16×16). Se puede observar que tienen cierta similitud con las funciones base de la transformada discreta del coseno

base con las que nos estamos quedando son las de alta frecuencia, atenuando las de baja frecuencia. Por lo tanto si hay ruido en la observación éste será amplificado dominando sobre la señal, que es lo que ocurre en la figura 5.1.d. De hecho uno de los métodos de regularización consiste en truncar la SVD eliminando los valores singulares pequeños [81, 82, 83].

5.3. Regularización de problemas inversos mal planteados

La regularización [42] es un método para construir soluciones aproximadas a problemas mal planteados en el que se construye un **operador de regularización**. Un operador

$$\mathcal{R}(g_d, \lambda) : G \longrightarrow F$$

dependiente de un parámetro λ y de la observación aproximada g_d se dice que es un operador de regularización para la ecuación 5.1 en una vecindad de g_o si cumple las siguientes propiedades:

1. Existe un $\delta_1 > 0$ tal que $\mathcal{R}(g_d, \lambda)$ está definido para todo λ y cualquier $g_\delta \in G$ para el cual $\rho_G(g_\delta, g_o) < \delta \leq \delta_1$
2. Existe una función $\lambda = \lambda(\delta)$ tal que para cualquier $\epsilon > 0$ existe un $\delta(\epsilon) \leq \delta_1$ que cumple que si $g_\delta \in G$ y $\rho_G(g_T, g_\delta) < \delta(\epsilon)$ entonces $\rho_F(f_\delta, f_o) < \epsilon$, donde $f_\rho = \mathcal{R}(g_\delta, \lambda(\delta))$.

La primera de las condiciones exige la existencia del operador de regularización entorno a g_o y la segunda exige la continuidad del operador de regularización.

Por tanto, para obtener una solución a un problema inverso mal planteado, primero se construye un operador de regularización y después se determina el *parámetro de regularización* λ.

Sea $F_0 \subset F$ una clase de soluciones admisibles (esta clase puede ser por ejemplo soluciones que pertenezcan al conjunto de las funciones de cuadrado integrable, al conjunto de funciones de m veces derivables, etc). Se pueden buscar soluciones aproximadas a 5.1 entre los elementos $f \in F_0$ que sean compatibles con los datos, es decir, tal que $\rho_G(Hf, g_d) \leq \delta$. Sea $F_0^\delta \subset F_0$ el conjunto de tales elementos compatibles con los datos. Si resulta que en la clase de funciones donde buscamos la solución el conjunto F_0^δ es vacío significa que se han buscado soluciones en una clase de funciones demasiado restrictiva (ya que ningún elemento de esta clase de funciones es compatible con los datos). Se puede pensar en tomar una secuencia de clases de soluciones $F_0 \subset F_1 \subset \cdots \subset F_n$ hasta obtener una clase que contenga elementos y que sean compatibles con los datos iniciales.

Si F_n es no vacía, entonces puede contener elementos (funciones, vectores, ... en función del problema tratado) que sean diferentes entre si. En tales casos se necesita un cierto principio para seleccionar soluciones compatibles con la observación, g_d. Tal criterio puede ser la elección de elementos de F_n que tengan mínima complejidad. La noción de complejidad de un elemento $f \in F_n$ puede ser formalizada, por ejemplo, definiendo funcionales de complejidad $\Omega(f)$ (ejemplos

de funcionales de complejidad son: la norma de la solución, la norma de la primera derivada de la solución, la norma de la segunda derivada de la solución,...).

La búsqueda de soluciones aproximadas de la ecuación $Hf = g_d$ entre los elementos más simples que sean compatibles con los datos iniciales conduce al problema de encontrar un elemento de F_n que minimiza $\Omega(f)$ sobre F_n^δ.

La figura 5.5 es una ilustración de estos conceptos utilizando el problema de la restauración. Para cada valor del parámetro de regularización se obtiene una solución. Del conjunto de soluciones posibles hay un subconjunto de soluciones aceptables (que se parecen a la imagen original) el resto o son demasiado ruidosas o son demasiado suaves.

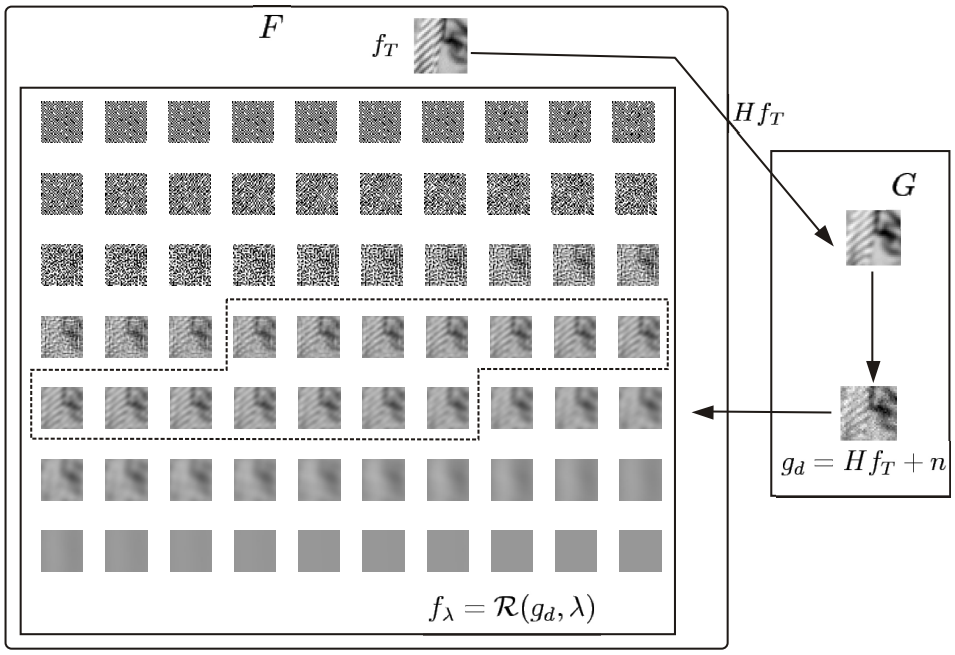


Fig. 5.5: Para cada valor del parámetro de regularización se obtiene una posible solución, f_λ. De este conjunto de soluciones sólo unas cuantas son aceptables (las que aparecen dentro de la línea punteada).

El problema se reduce a encontrar un elemento f_λ en el conjunto $F_n^\delta \bigcap F_\Omega$ que minimiza el funcional $M^\lambda(f, g_d)$

$$f_\lambda = \arg\min_f M^\lambda(f, g_d) = \arg\min_f \left(\rho_G^2(Hf, \tilde{g}) + \lambda^2 \Omega(f)\right) \tag{5.8}$$

El elemento f_λ que minimiza $M^\lambda(f, g_d)$ puede ser visto como el resultado de

aplicar sobre g_d un operador que depende de λ, es decir, $f = R(g_d, \lambda)$ [84].

5.4. Regularización de la restauración

En el caso de la restauración, la ecuación 5.8 se puede expresar como

$$i(\mathbf{x})' = \arg\min_i \big(f_1(i(\mathbf{x}), i_d(\mathbf{x})) + \lambda^2 f_2(i(\mathbf{x}))\big) \tag{5.9}$$

donde el funcional $f_1(\cdot)$ debe medir la desviación de la solución respecto de los datos observados y el funcional $f_2(\cdot)$ debe medir la presencia de características no deseadas en la solución, es decir, $f_2(\cdot)$ es el funcional de complejidad comentado en la sección anterior. Minimizar esta ecuación implica encontrar una solución que sea compatible con los datos observados y que posea en la menor medida posible características no deseadas.

Si se utiliza la norma 2 en ambos funcionales, y se utiliza un operador de penalización en el funcional $f_2(\cdot)$ esta ecuación se puede expresar como:

$$i' = \arg\min_i \|i(x,y) \star h(x,y) - i_d(x,y))\|^2 + \lambda^2 \|p(i(x,y)))\|^2$$

donde $p(\cdot)$ es el operador de penalización, que debe extraer las características no deseadas en la solución. Se puede ver [80] que el problema es formalmente equivalente al ajuste por mínimos cuadrados con restricciones. Si tomamos una aproximación local se puede considerar que tanto la respuesta impulsional $h(x,y)$ como el operador $p(\cdot)$ son espacialmente invariantes, la solución en una región x_0 de un dominio de Fourier por bloques es [85]:

$$c'(x_0, f) = \frac{H^*(x_0, f)}{|H(x_0, f)|^2 + \frac{\lambda(x_0)^2}{|P(x_0,f)^{-1}|^2}}\, c_d(x_0, f) \tag{5.10}$$

La ventaja de realizar una aproximación local es que, aun teniendo señales y degradaciones que no sean globalmente estacionarias, podemos aplicar la aproximación de invariancia espacial y la adaptar tanto el operador a posibles cambios en diferentes regiones espaciales como el parámetro de regularización a ruidos cuya energía es espacialmente variante.

Para aportar cierta intuición sobre el significado de los parámetros presentes en el modelo de regularización, vamos a establecer una comparación con el resultado del filtro óptimo de Wiener [85]. En la solución de Wiener, el espectro de la imagen restaurada es:

$$c'(x_0, f) = \frac{H^*(x_0, f)}{|H(x_0, f)|^2 + \frac{S_{nn}(x_0,f)}{S_{i_o i_o}(x_0,f)}} \, c_d(x_0, f) \tag{5.11}$$

donde $S_{nn}(x_0, f)$ y $S_{i_o i_o}(x_0, f)$ son la densidad espectral de potencia del ruido y la señal respectivamente. Obviamente, las ecuaciones 5.10 y 5.11 son formalmente equivalentes. Identificando términos se observa que el parámetro λ^2 está relacionado con la energía del ruido y el operador de penalización, con la densidad espectral de potencia de la señal.

Puesto que se está utilizando una aproximación local (la imagen se divide en regiones y cada región se restaura por separado) es posible que se produzcan discontinuidades entre bloques vecinos. Para suavizar las soluciones entre bloques vecinos se han tomado regiones que presentan solapamiento de forma que la solución en el solapamiento será la media de las soluciones vecinas.

Elementos relevantes en la restauración mediante regularización

De la ecuación 5.9 podemos concluir que la solución depende de:

1. Elección de los funcionales $f_1(\cdot)$ y $f_2(\cdot)$.
2. Selección de un operador de penalización en $f_2(\cdot)$.
3. Selección de λ ó estimación de $S_{nn}(x_0, f)$.
4. Conocimiento de la respuesta impulsional en el dominio espacial o frecuencial (h ó H).

Respecto del primer punto hemos de hacer notar que existen diferentes posibilidades en la literatura. El funcional f_2 puede utilizar la norma 1 en lugar de la norma 2. La motivación es que la norma 1 penaliza menos las discontinuidades que la norma 2 (cuando se combina la norma 1 con el operador primera derivada se está ante lo que se denomina como *Total Variation Restoration* [86, 87]).

También hemos de decir que los puntos 2 y 3 están estrechamente relacionados, pues la elección de un operador adecuado relajará la exigencia de una buena estimación de λ. Claramente, un operador que discrimine convenientemente la señal del ruido restará importancia al grado de conocimiento del ruido.

Por último, es evidente que en una imagen degradada real no va a conocerse a priori el proceso de suavizado, con lo cual, resulta de la mayor importancia un método eficaz para la estimación de la respuesta impulsional de la degradación a partir de la imagen observada (deconvolución ciega). En este trabajo se asume que la respuesta impulsional de la degradación es conocida.

En este trabajo nos centraremos en la selección del parámetro de regularización y en los operadores de penalización.

5.5. Selección del parámetro de regularización

El criterio de la L-curva [88, 43, 83] está basado en una gráfica del término de penalización $\|Pi_\lambda\|$ frente a la norma residual $\|Hi_\lambda - i_d\|$ para diferentes valores de λ. La curva resultante tiene una parte vertical y una parte horizontal y una esquina que separa estas dos regiones y de ahí su nombre.

La parte horizontal corresponde a soluciones i_λ que son suaves y la parte vertical corresponde a soluciones que son muy ruidosas. Por lo tanto la esquina de la L-curva que separa la parte vertical y la horizontal debe corresponder a un buen balance entre minimizar el error de regularización y el error de perturbación en la solución. Hay que decir que este método está basado en una aproximación heurística y no hay resultados sobre su convergencia.

La definición de esquina de la L-curva propuesta en [43] es el punto de la L-curva en escala logarítmica con máxima curvatura. Si $\rho = \log \|Hi_\lambda - i_d\|$ y $\eta = \log \|Pi_\lambda\|$ entonces la curvatura como una función del parámetro de regularización viene dada por:

$$\mathcal{K}(\lambda) = \frac{\rho'\eta'' - \rho''\eta'}{((\rho')^2 + (\eta')^2)^{\frac{3}{2}}} \tag{5.12}$$

Utilizando las funciones contenidas en la *Regularization Toolbox* desarrolladas por P. C. Hansen[2] para restaurar una imagen se obtiene una solución en la que hay mucha variabilidad entre los bloques (coexisten en la solución bloques con ruido y bloques demasiado suaves) tal y como se puede observar en el resultado mostrado en la figura 5.6.a.

5.5.1. Propuesta de modificación del criterio de la L-curva

Con el fin de evitar esta variabilidad en la solución, se ha realizado una modificación del criterio para la obtención de la esquina de la L-curva. El método consiste en realizar una interpolación para obtener una curva con más puntos, obtener el ángulo que forma la recta tangente en cada punto interpolado y tomar como esquina de la L-curva el punto en el que se produce mayor variación en el ángulo. El proceso es descrito en la figura 5.7.

Con este nuevo método las soluciones para cada bloque no tienen tanta variabilidad tal y como se puede ver en la figura 5.6.b.

[2] Disponible en la dirección http://www.imm.dtu.dk/pch/Regutools/index.html

a) b)

Fig. 5.6: a) Soluciones obtenidas para una imagen degradada con un filtro pasa baja con frecuencia de corte $f_c = 8$ cpd y ruido blanco gaussiano de varianza $\sigma^2 = 50$. a) Solución obtenida utilizando las funciones de la *Regularization Toolbox* y utilizando un operador de penalización basado en la segunda derivada. b)Solución obtenida utilizando el criterio modificado de obtención del parámetro de regularización.

5.6. Operadores de penalización y características de la señal

Como ya se ha comentado se debe elegir un operador para regularizar. Una vez se dispone del operador y de un método de selección del parámetro de regularización es posible obtener la solución utilizando la ecuación 5.10. En esta sección se revisarán algunos operadores de penalización clásicos y se propondrá el operador adaptativo basado en la respuesta no lineal del sistema visual humano y se comparará con otras aproximaciones espacialmente adaptativas basadas en el ajuste de modelos autorregresivos. Se mostrarán todas las soluciones obtenidas para dos bloques degradados lo cual motivará la búsqueda de medidas para evaluar los diferentes operadores.

5.6.1. Operadores de penalización clásicos

Uno de los operadores de penalización no adaptativo más utilizado es el basado en la segunda derivada [80, 79, 89] que denotaremos como L_2. La característica no deseada que mide este operador es la presencia en la solución de componentes de alta frecuencia. Esto es lógico (compatible con modelos sencillos de estadística de las imágenes naturales) ya que el espectro de las imágenes naturales decae con la frecuencia según $\frac{1}{f}$ [90, 26, 23] por lo tanto hay que limitar la presencia de altas frecuencias en la solución. El problema de de L_2 es que no está formulado

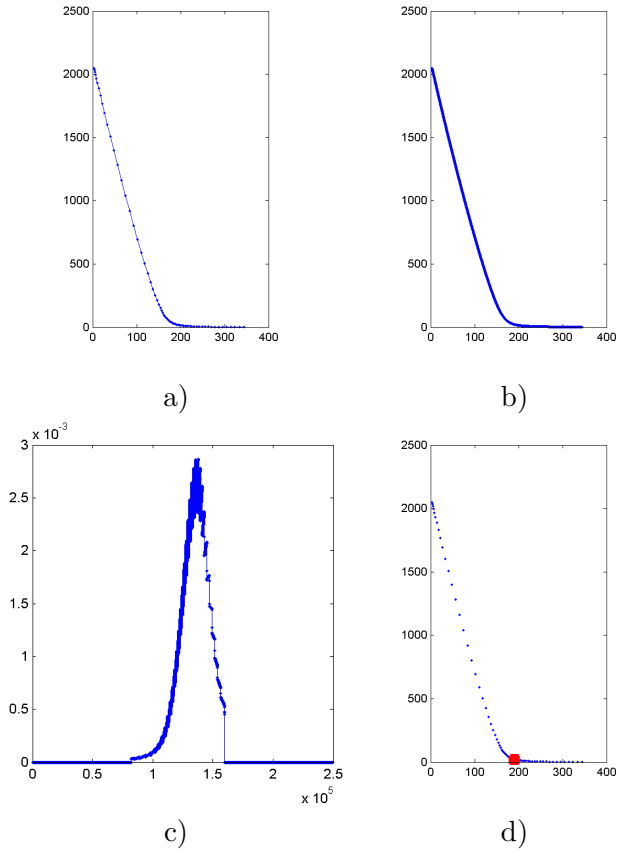


Fig. 5.7: a) Puntos donde se ha obtenido la L-curva. b) Curva con mayor densidad de puntos utilizando interpolación cúbica. c) Variación del ángulo de la recta tangente. c) Punto de la L-curva que proporciona el parámetro de regularización óptimo correspondiente a la variación máxima del ángulo.

en Fourier, es fijo y no se podría controlar por ejemplo la distancia a la que se está viendo la imagen (que modifica lo que se considera altas o bajas frecuencias).

El sistema visual humano actua como un filtro de frecuencias espaciales, el ojo se comporta como un filtro pasa-baja que no puede percibir frecuencias espaciales más allá de un determinado límite. El sistema visual percibe mejor las frecuencias espaciales intermedias que las bajas y las altas frecuencias [11]. Por lo tanto es posible utilizar la inversa de la CSF como operador de penalización [85]. De esta forma también se consigue penalizar las altas frecuencias pero con un filtro que está definido en unidades con significado físico (ciclos por grado) en lugar del operador L_2 definido directamente en el dominio de la imagen.

Los modelos autorregresivos, AR, se pueden utilizar para definir operadores

adaptativos y han sido utilizados en el contexto de la restauración [80]. En un modelo autorregresivo, el valor de la señal en un punto, $s(i,j)$, se obtiene como una combinación lineal de los valores de la señal en una vecindad del punto, $\mathcal{N}(i,j)$, más un término de error

$$s(i,j) = \sum_{(l,m)\in\mathcal{N}(i,j)} a(l,m)s(l,m) + e(i,j)$$

En imágenes tiene sentido utilizar este tipo de modelos ya que las imágenes naturales presentan alta correlación entre puntos vecinos, por lo tanto es posible predecir el valor de un punto a partir del valor de los puntos vecinos. La figura 5 de la **Trabajo E** muestra la forma de las vecindades consideradas (una vecindad con 4 coeficientes, AR4, una vecindad con 8 coeficientes, AR8, y una vecindad con 12 coeficientes, AR12).

Una vez seleccionado el modelo AR, los coeficientes se ajustan de forma independiente para cada bloque mediante mínimos cuadrados, adaptando de esta forma el operador a las características locales. El operador, que hemos definido como la inversa de la densidad espectral de potencia, se obtiene a partir de los coeficientes AR [91]

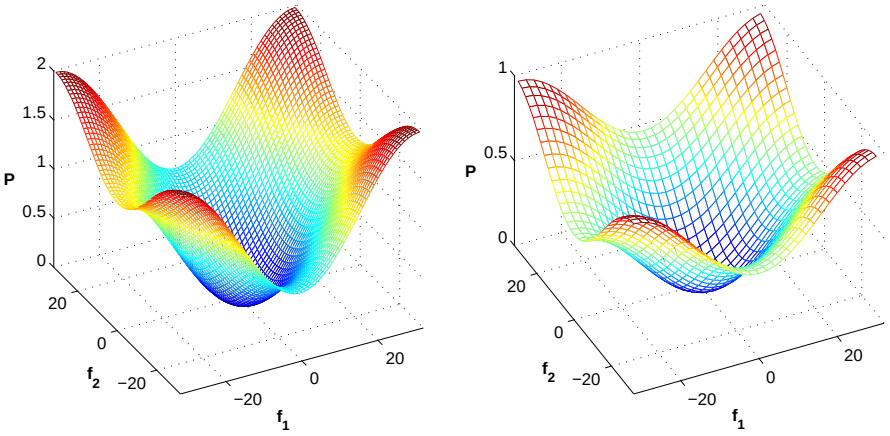
$$P(w_1,w_2) = \frac{\sigma^2}{|1+\sum_{n_1}\sum_{n_2} a(n_1,n_2)e^{-jw_1n_1}e^{-jw_2n_2}|^2}$$

Es interesante resaltar que si se ajustan los coeficientes de un modelo autorregresivo de orden 4 con la estructura mostrada en la figura 5 de la **Trabajo E**, utilizando como conjunto de entrenamiento imágenes naturales extraídas por ejemplo de la base de imágenes de van Hateren [55], se obtiene un operador de penalización que es casi idéntico al operador basado en la segunda derivada, tal y como se muestra en la figura 5.8.

5.7. Operador de penalización propuesto

Como ya se comentó en el capítulo anterior, la hipótesis de la codificación eficiente establece que el sistema visual humano está ajustado a la estadística de las imágenes naturales, obteniendo las características relevantes de la señal. Si disponemos de un modelo de respuesta podríamos definir un operador de penalización como la inversa de la respuesta, ya que no penalizaríamos la presencia de características que el sistema visual considera relevantes (véase la figura de la **Trabajo E** para una ilustración de este hecho para un bloque particular).

De este modo, si $R(c_d(x_o,f))$ representa la respuesta ante un bloque de la imagen degradada, el operador vendría dado por



a) b)

Fig. 5.8: a) Operador (en Fourier) resultante del ajuste de un modelo AR4 utilizando un conjunto de imágenes naturales. b) Operador L_2 (en Fourier). Como se puede observar, el operador resultante del ajuste del modelo AR4 es muy similar a L_2.

$$P(x_o, f) = \frac{1}{R(c_d(x_o, f))}$$

En este trabajo se ha utilizado la respuesta no lineal con interacción entre coeficientes presentada en la sección 2.5. Esquemas estadísticos recientes explotan también la existencia de relaciones entre los coeficientes del dominio wavelet [92].

5.7.1. *Ventajas del operador perceptual*

Las figuras 5.9- 5.14 muestran todas las soluciones que se obtienen cuando se varía el parámetro de regularización al utilizar diferentes operadores de penalización, sobre una imagen degradada utilizando un filtro basa baja con frecuencia de corte $f_c = 22$ cpd y ruido blanco gaussiano de varianza $\sigma^2 = 200$.

Como se puede observar en estas figuras, el rango de soluciones *aceptables* depende del operador utilizado. La figura 5.15 muestra la distancia entre las soluciones y el bloque original en función del parámetro de regularización. Como se puede observar, en ambos casos el operador que proporciona un rango de soluciones con menor error es el basado en la respuesta no lineal del SVH, y por lo tanto la selección del parámetro de regularización no es tan crítica como con los otros operadores.

Esta observación motiva la búsqueda de medidas que indiquen por un lado la

sensibilidad de la solución en función del parámetro de regularización al utilizar un determinado operador de penalización y la robustez de los diferentes operadores de penalización frente a la degradación.

5.7.2. Sensibilidad

Como se ha mostrado en el apartado anterior, la sensibilidad de la solución respecto al parámetro de regularización depende de la capacidad del operador de capturar las propiedades de la señal: la solución es más sensible si pequeñas variaciones en el parámetro de regularización dan lugar a cambios considerables en la calidad de la solución.

Cuando se utiliza un operador concreto, P, se obtiene una determinada solución, $\hat{i}_P(\lambda)$, para cada valor de λ. Un método particular para la selección del parámetro de regularización, como por ejemplo la L-curva proporciona un valor particular λ_L, que da lugar a una solución particular $\hat{i}_P(\lambda_L)$. Asumiendo que la imagen original, i_o, es conocida, y que se está utilizando una determinada medida de distorsión $d(\cdot,\cdot)$, el valor óptimo (no disponible en una situación real ya que implica el conocimiento de i_o) de λ vendría dado por

$$\lambda_{opt} = \arg\min_{\lambda} D_P(\lambda)$$

donde $D_P(\lambda) = d(i_o, \hat{i}_P(\lambda))$. La sensibilidad de la solución respecto de λ cuando se utiliza un determinado operador P, puede ser definida como el incremento de distorsión debido a la selección de λ_L en lugar del valor óptimo λ_{opt}, es decir

$$S(P) = \frac{\Delta D_P}{\Delta \lambda} = \frac{|D_P(\lambda_{opt}) - D_P(\lambda_L)|}{|\lambda_{opt} - \lambda_L|}$$

El significado de esta medida de sensibilidad se ilustra en la figura 5.16

En la figura 7 de la **Trabajo E** se muestran los resultados obtenidos para la imagen Barbara utilizando los diferentes operadores de penalización y utilizando como medida de distorsión el MSE. Como se puede observar esta medida refleja correctamente los resultados mostrados en el apartado anterior, es decir, el operador perceptual tiene una sensibilidad menor que el operador AR12 y el operador L2.

5.7.3. Robustez

Deseamos que la medida de robustez de un operador refleje la habilidad de un operador adaptativo de preservar los mismos componentes de la señal degradada

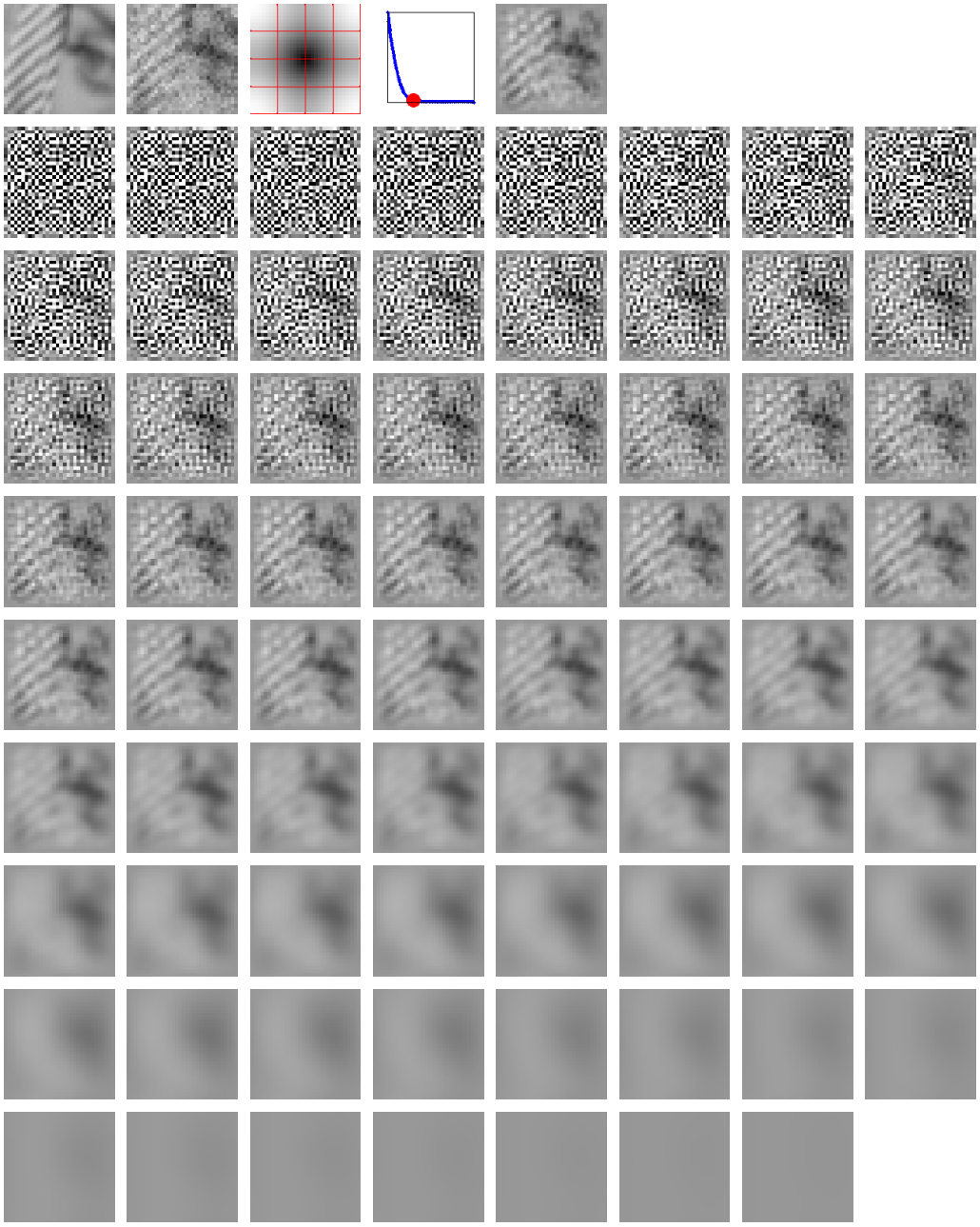


Fig. 5.9: Análisis de los resultados para un bloque utilizando un operador de penalización basado en la segunda derivada. En la primer fila se muestra el bloque original, el bloque degradado, el modulo de la respuesta en frecuencia del operador de penalización, la L-curva y posición en la curva correspondiente al parámetro de regularización óptimo y finalmente el bloque restaurado. El resto de las filas muestra la solución para cada valor del parámetro de regularización. Este operador es igual para todos los bloques independientemente de su contenido.

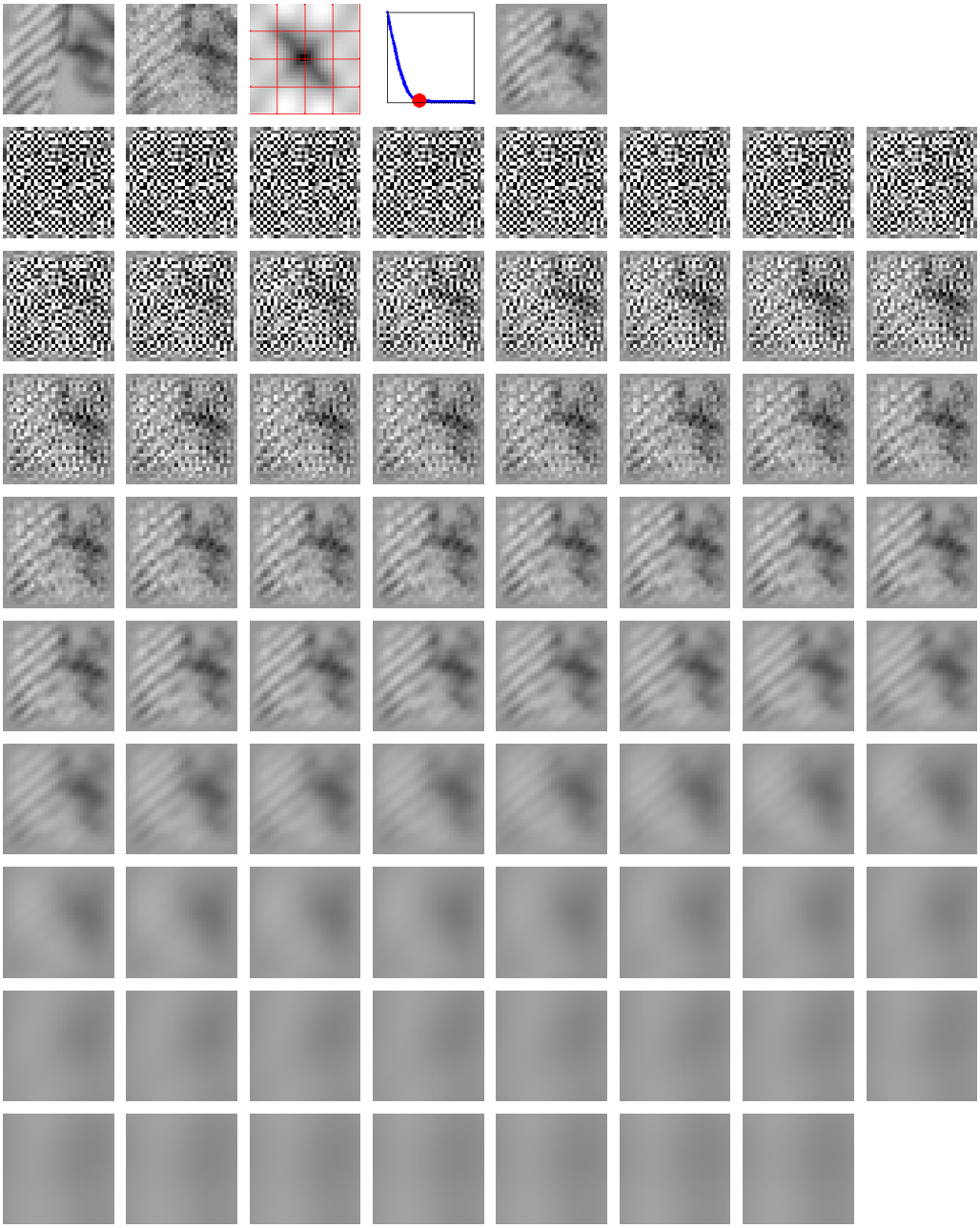


Fig. 5.10: Análisis de los resultados para un bloque utilizando un operador de penalización basado en un modelo autorregresivo de orden 12. En la primer fila se muestra el bloque original, el bloque degradado, el modulo de la respuesta en frecuencia del operador de penalización, la L-curva y posición en la curva correspondiente al parámetro de regularización óptimo y finalmente el bloque restaurado. El resto de las filas muestra la solución para cada valor del parámetro de regularización. Como se puede observar, el operador resultante no penaliza la presencia de las rayas oblicuas.

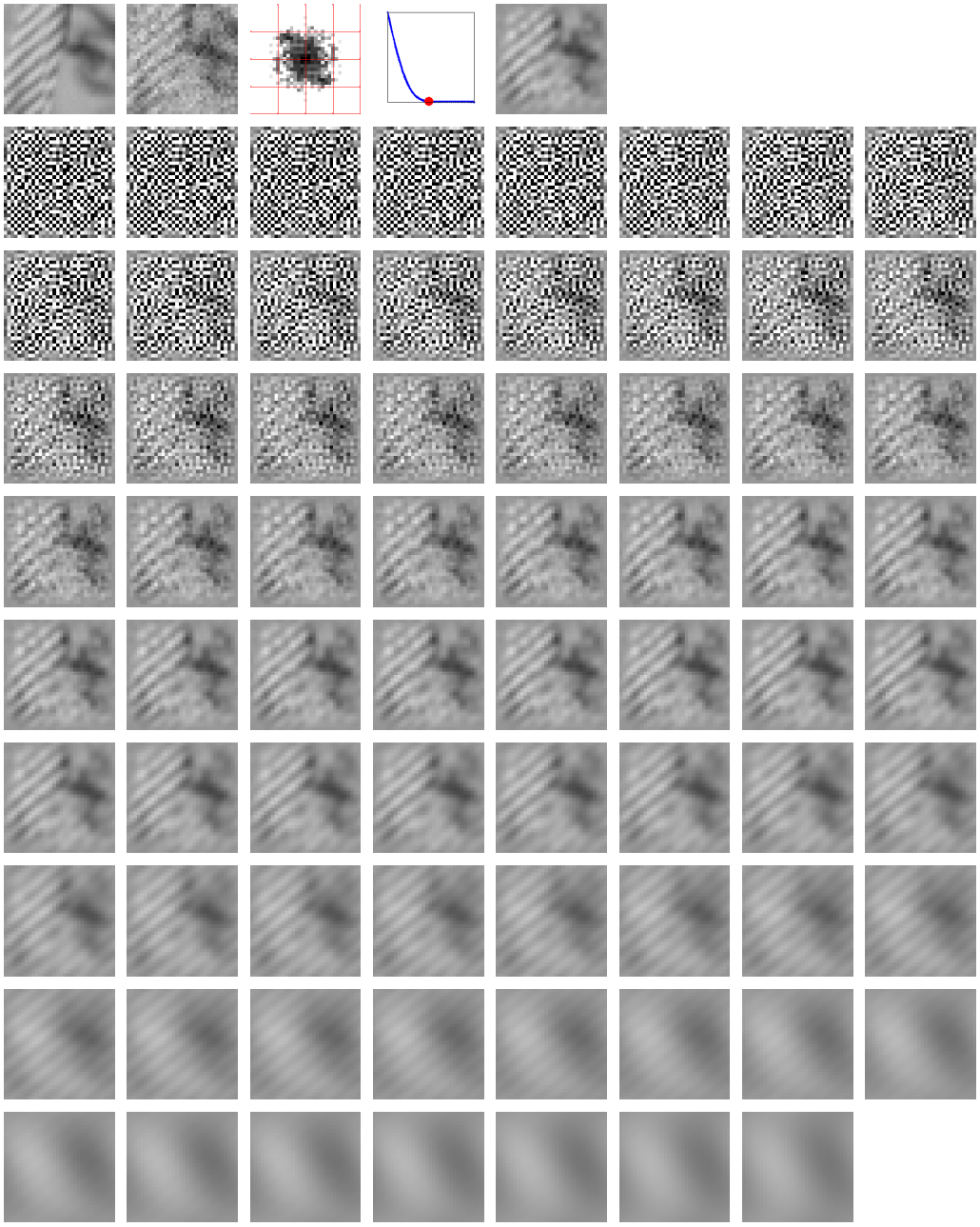


Fig. 5.11: Análisis de los resultados para un bloque utilizando un operador de penalización basado en en el modelo no lineal de respuesta del sistema visual humano. En la primer fila se muestra el bloque original, el bloque degradado, el modulo de la respuesta en frecuencia del operador de penalización, la L-curva y posición en la curva correspondiente al parámetro de regularización óptimo y finalmente el bloque restaurado. El resto de las filas muestra la solución para cada valor del parámetro de regularización. Como se puede observar, el operador resultante no penaliza la presencia de las rayas oblicuas.

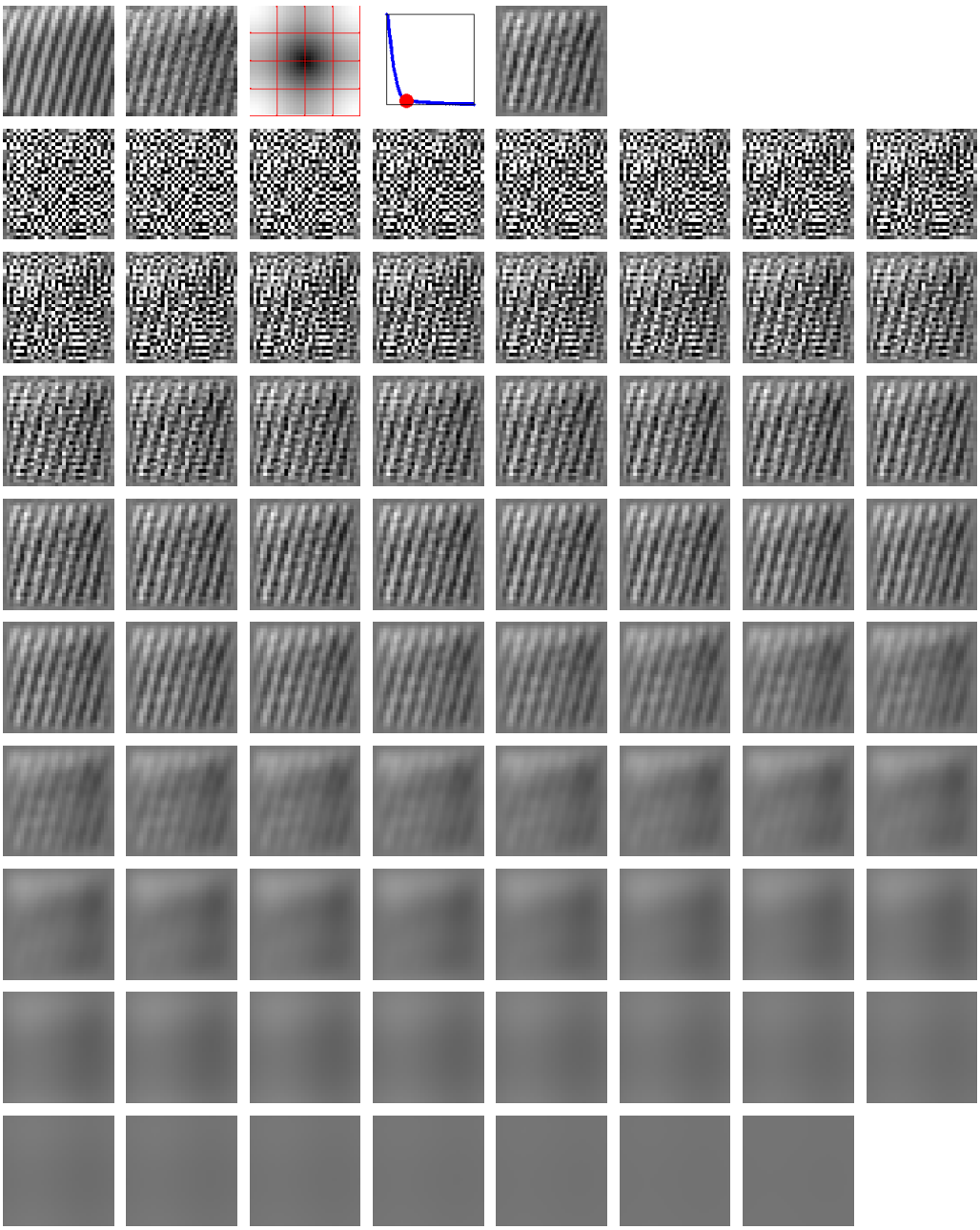


Fig. 5.12: Análisis de los resultados para un bloque utilizando un operador de penalización basado en la segunda derivada. En la primer fila se muestra el bloque original, el bloque degradado, el modulo de la respuesta en frecuencia del operador de penalización, la L-curva y posición en la curva correspondiente al parámetro de regularización óptimo y finalmente el bloque restaurado. El resto de las filas muestra la solución para cada valor del parámetro de regularización.

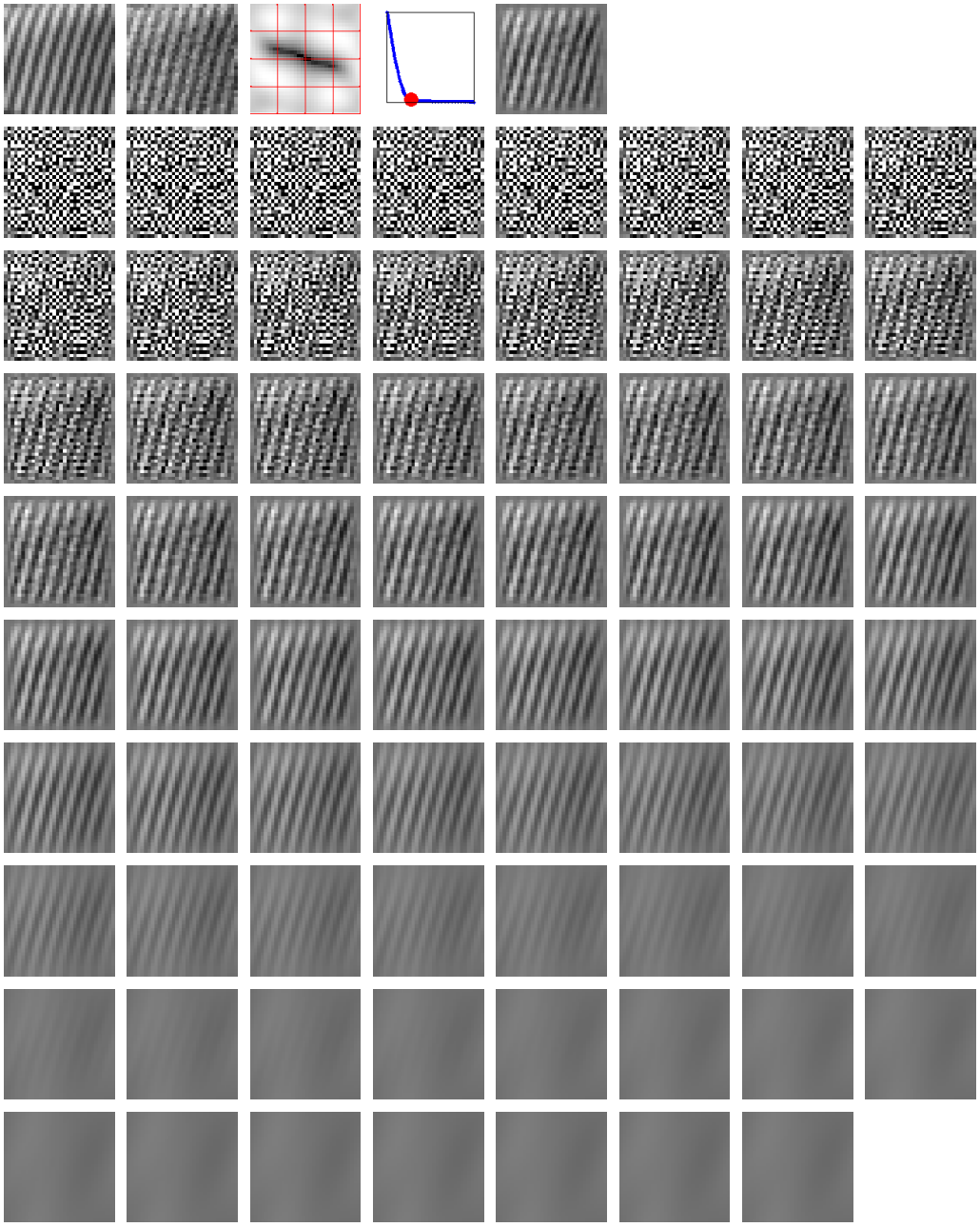


Fig. 5.13: Análisis de los resultados para un bloque utilizando un operador de penalización basado en un modelo autorregresivo de orden 12. En la primer fila se muestra el bloque original, el bloque degradado, el modulo de la respuesta en frecuencia del operador de penalización, la L-curva y posición en la curva correspondiente al parámetro de regularización óptimo y finalmente el bloque restaurado. El resto de las filas muestra la solución para cada valor del parámetro de regularización.

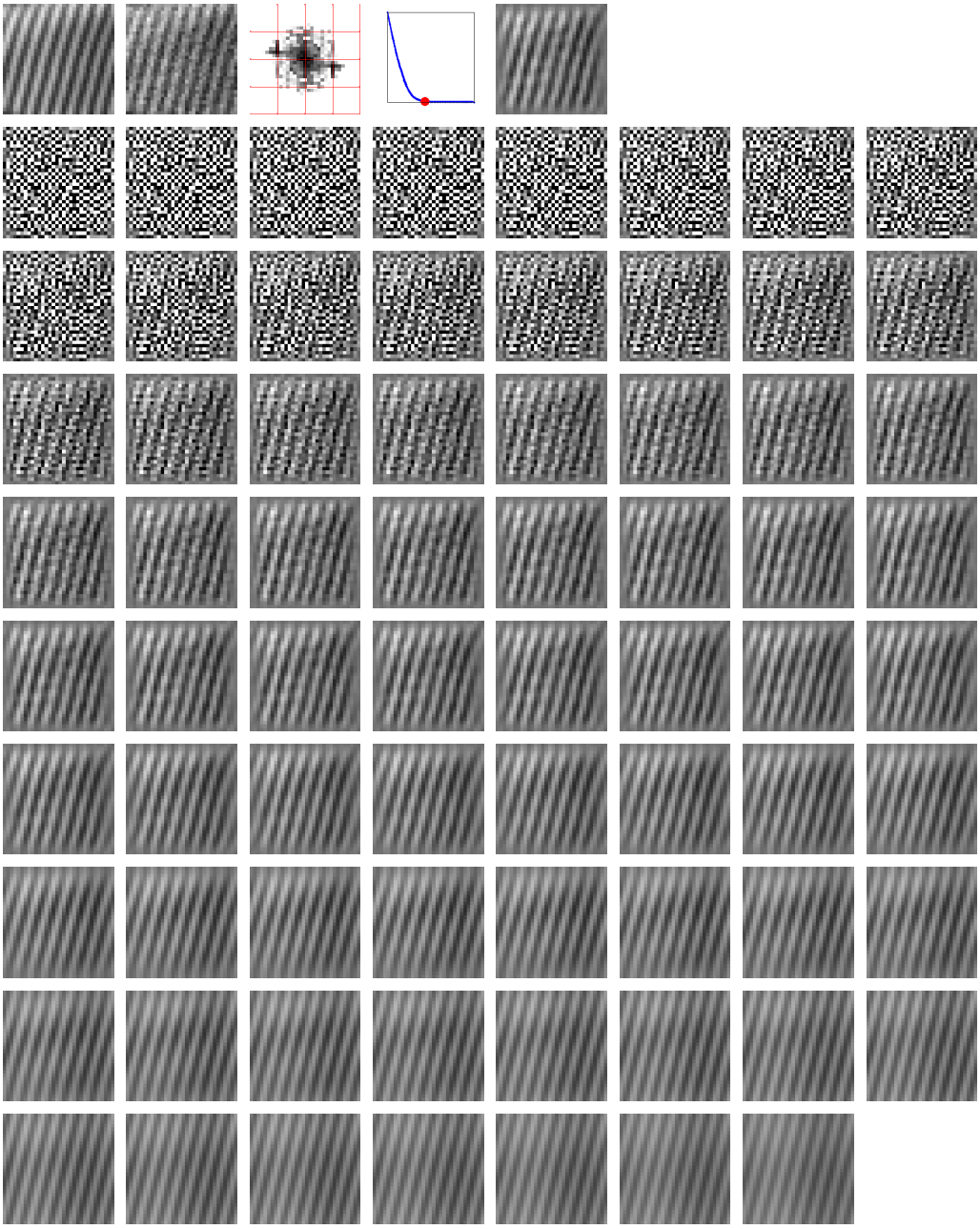


Fig. 5.14: Análisis de los resultados para un bloque utilizando un operador de penalización basado en en el modelo no lineal de respuesta del sistema visual humano. En la primer fila se muestra el bloque original, el bloque degradado, el modulo de la respuesta en frecuencia del operador de penalización, la L-curva y posición en la curva correspondiente al parámetro de regularización óptimo y finalmente el bloque restaurado. El resto de las filas muestra la solución para cada valor del parámetro de regularización.

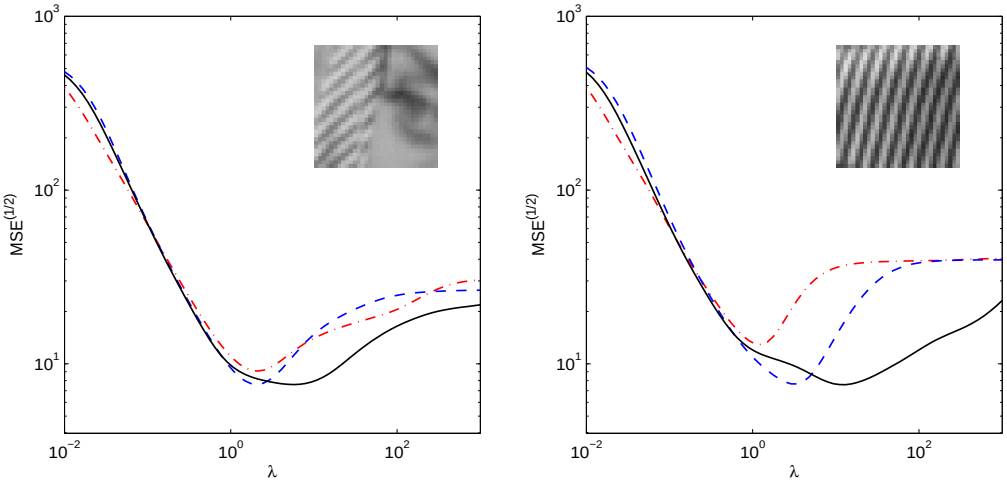


Fig. 5.15: Distancia ($\text{MSE}^{\frac{1}{2}}$) entre el bloque original y las diferentes soluciones obtenidas utilizando los diferentes métodos para cada valor de λ. Línea continua: con el operador propuesto, línea discontinua: con el operador basado en el modelo AR(12) y en línea discontinua punteada: con el operador basado en la segunda derivada.

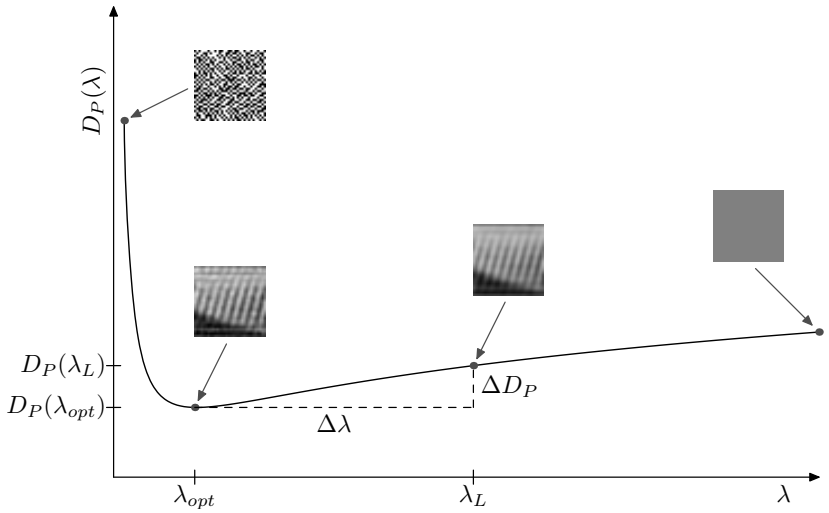


Fig. 5.16: Ilustración de las variables utilizadas para definir la medida de sensibilidad.

que serían preservados con un operador ideal obtenido a partir de la señal original (sin degradar).

Si se dispone de la imagen original, i_o, el operador adaptativo P_o estimado

utilizando esta imagen se puede comparar con el operador adaptativo P obtenido a partir de la imagen degradada. El MSE entre los dos operadores puede ser utilizado para medir la similitud entre estos dos operadores. Por lo tanto, definimos la robustez de un operador adaptativo como

$$R(P) = \text{MSE}(P, P_o)$$

En la figura 8 de la **Trabajo E** se muestran los resultados obtenidos para la imagen Barbara utilizando los diferentes operadores de penalización adaptativos. Puesto que el operador perceptual propuesto captura mejor la complejidad de la señal, es más robusto frente a la presencia de la degradación.

5.7.4. *Restauración de imágenes*

Ruido estacionario

Las tablas I,II y III y las figuras 10,11 y 12 de la **Trabajo E** muestran los resultados obtenidos con los diferentes operadores en la restauración de imágenes en las que se ha simulado la degradación. El rango de las degradaciones generadas va desde casi exclusivamente emborronamiento hasta casi exclusivamente eliminación de ruido.

La tabla IV y las figuras 13, 14 y 15 de la **Trabajo E** muestran los resultados obtenidos en eliminación de ruido con imágenes tomadas bajo diferentes condiciones de iluminación con una cámara SONY DSC P8. En este caso comparamos los resultados obtenidos utilizando el operador propuesto con los resultados obtenidos utilizando un algoritmo estándar de eliminación de ruido [93].

Ruido de energía espacialmente variante

En este apartado se presentan resultados (que no están incluidos en las publicaciones) de eliminación de ruido cuya energía es espacialmente variante. La figura 5.17.a muestra la imagen degradada resultante de añadir el ruido que se muestra en la figura 5.17.c a la imagen Barbara. El ruido tiene una varianza en el rango $\sigma^2 \in [0, 300]$ de izquierda a derecha. La figura 5.17.b muestra la imagen restaurada utilizando el operador perceptual adaptativo y la figura 5.17.d muestra la resta entre la imagen restaurada y la degradada, esta resta debería coincidir con el ruido. La figura 5.17.e muestra la varianza del ruido real en función de la columna y la figura 5.17.f muestra cual es la varianza del ruido estimado. Hay que resaltar que el hecho de que el método se adapte a ruidos de energía espacialmente variante viene determinado por la estimación en cada bloque del parámetro de regularización.

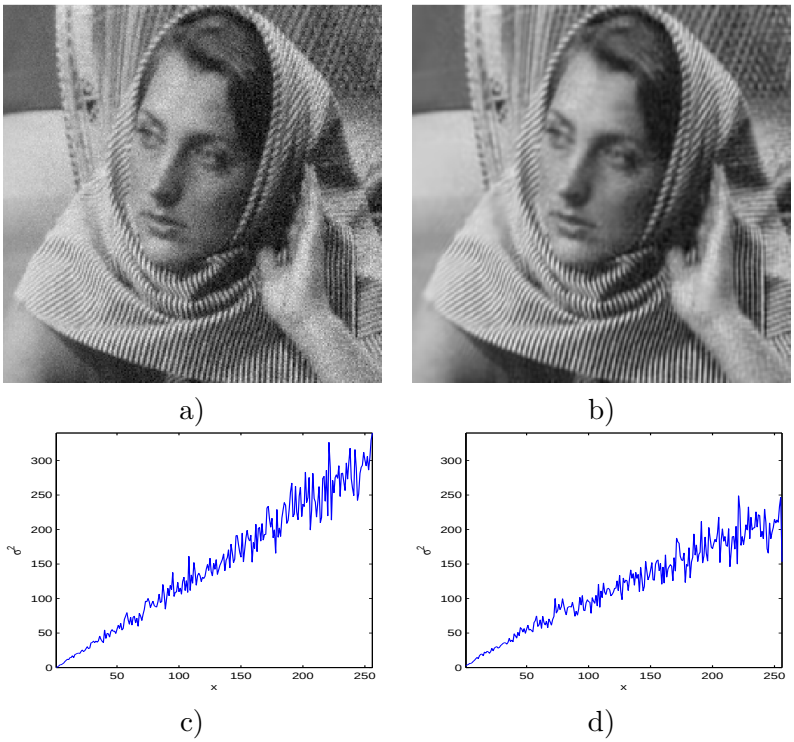


Fig. 5.17: a) Imagen degradada. b) Imagen restaurada utilizando el operador perceptual. c) Varianza del ruido aplicado. d) Varianza del ruido estimado (resultado de realizar la resta entre la imagen degradada y la imagen restaurada).

6. CONCLUSIONES Y TRABAJO FUTURO

La hipótesis de la codificación eficiente establece que debe existir una relación entre la estadística de los estímulos y la organización de los sistemas sensoriales biológicos. En esta Tesis hemos mostrado resultados adicionales (y más generales) en favor de esta hipótesis, y hemos aplicado las no linealidades de la respuesta resultante como alternativa a modelos estadísticos explícitos en dos aplicaciones de procesado de imágenes: estimación del movimiento en secuencias de imágenes y restauración de imágenes degradadas.

En cuanto a la relación entre la respuesta de los sensores de V1 y la estadística de las imágenes naturales, durante la última década se había logrado reproducir la forma de los campos receptivos de V1 y ciertas propiedades de las no linealidades asumiendo una forma funcional concreta (la normalización divisiva). En esta Tesis hemos mostrado que las propiedades más importantes de la parte no lineal de dicha respuesta (saturación y enmascaramiento cruzado) emergen directamente de los datos sin necesidad de asumir ninguna forma funcional. Para ello hemos propuesto una nueva solución para el ICA no lineal basado en una aproximación diferencial, en la que a partir de ICAs lineales locales se llega a una representación global no lineal.

Los esquemas estándar de codificación de video están basados en un módulo de compensación de movimiento y en un módulo de cuantización del error de predicción. La literatura anterior había propuesto la introducción de información perceptual en los dos módulos, pero no se había estudiado sistemáticamente el efecto de estas mejoras ni en la calidad subjetiva de la señal decodificada, ni en la utilidad del flujo óptico para tareas de más alto nivel. En esta Tesis se ha mostrado que las mejoras en el cuantizador, basadas en la introducción del modelo no lineal de percepción, son más importantes en cuanto a la calidad de la secuencia reconstruida que las mejoras debidas a una óptima distribución de bits entre la estimación de movimiento y la cuantización del error. No obstante, se ha mostrado que el criterio de refinamiento del flujo óptico basado en la entropía perceptual produce una estimación del movimiento más robusta ya que produce mejores segmentaciones cuando se utiliza como entrada a un algoritmo de estimación de VOPs.

En la restauración de imágenes degradadas en el contexto de la teoría de la regularización, el operador de penalización debe contener información sobre las

características rechazables de la solución. En la literatura se habían propuesto diferentes operadores basados en las características espectrales de las imágenes naturales. En esta Tesis se han definido operadores de penalización basados en el modelo de percepción no lineal con interacción entre coeficientes y se han comparado con operadores adaptativos basados en modelos autorregresivos. Se ha mostrado que la consideración de la interacción entre coeficientes tiene consecuencias positivas: los operadores son más robustos, las soluciones son menos sensibles al parámetro de regularización y se comportan mejor en un amplio rango de situaciones (desde situaciones cercanas al desemborronamiento hasta situaciones cercanas a la eliminación de ruido). En la eliminación de ruido en imágenes con degradaciones reales se ha mostrado que los operadores propuestos distinguen mejor entre el ruido y las estructuras relevantes de la imagen (que son penalizadas por un filtro de Wiener). Estos resultados sugieren que estos operadores, que no requieren conocimiento estadístico a priori sobre la imagen, pueden ser utilizados como alternativa a los operadores que se basan en la estimación del espectro de la imagen.

Como trabajo futuro es posible la aplicación del ICA no lineal sobre estímulos de mayor dimensionalidad, de esta forma se tendría mayor resolución frecuencial con el fin de comprobar si se puede dar cuenta del efecto de la frecuencia en el enmascaramiento cruzado. En cuanto a la restauración los resultados se pueden extender utilizando una norma cercana a la unidad en el funcional de penalización junto con los operadores propuestos ya que esta norma no impide la presencia de bordes en la solución. Esto sería una generalización de los trabajos de *Total Variation*, en los que se utiliza un operador primera derivada. En cuanto a las aplicaciones de compresión de imágenes estamos trabajando en la introducción de modelos no lineales en dos tipos de esquemas de codificación: los basados en transformadas wavelet, y otros más novedosos que utilizan máquinas de vectores soporte.

BIBLIOGRAFíA

[1] P. Lennie, "Single units and visual cortical organization," *Perception*, vol. 27, pp. 889–935, 1998.

[2] D. H. Hubel and T. N. Weisel, "Receptive fields, binocual interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, vol. 160, pp. 106–154, 1962.

[3] D. H. Hubel, "Evolution of ideas on the primary visual cortex, 1955-1978: a biased historial account," Nobel lecture, December 1981.

[4] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience*, vol. 9, pp. 181–198, 1992.

[5] A. Watson and J. Solomon, "A model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America A*, vol. 14, pp. 2379–2391, 1997.

[6] S. Marcelja, "Mathematical description of the response of simple cortical cells," *Journal of the Optical Society of America*, vol. 70, no. 11, pp. 1297–1300, 1980.

[7] J. Daugman, "Two-dimensional spectral analysis of cortical recptive field profiles," *Vision Research*, vol. 20, pp. 847–856, 1980.

[8] A. Watson, "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images*, ser. Springer Series on Information Sciences, O. Braddick and A. Sleigh, Eds., vol. 11. Berlin: Springer Verlag, 1983, pp. 100–114.

[9] ——, "The cortex transform: Rapid computation of simulated neural images," *Computer Vision, Graphics and Image Processing*, vol. 39, pp. 311–327, 1987.

[10] E. Simoncelli and E. Adelson, *Subband Image Coding*. Norwell, MA: Kluwer Academic Publishers, 1990, ch. Subband Transforms, pp. 143–192.

[11] F. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, pp. 551–566, 1968.

[12] G. Legge and J. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, vol. 70, pp. 1458–1471, 1980.

[13] G. Legge, "A power law for contrast discrimination," *Vision Research*, vol. 18, pp. 68–91, 1981.

[14] M. Carandini and D. Heeger, "Summation and division by neurons in visual cortex," *Science*, vol. 264, no. 5163, pp. 1333–6, 1994.

[15] M. Carandini, D. J. Heeger, and A. Movshon, "Linearity and normalization in simple cells of the macaque primary visual cortex," *Journal of Neuroscience*, vol. 17, no. 21, pp. 8621–8644, 1997.

[16] R. Linsker, "Perceptual neural organization: some approaches based on network models and information theory," *Annual Review of Neuroscience*, vol. 13, pp. 257–281, 1990.

[17] ——, "From basic network principles to neural architecture: emergence of orientation-selective cells," *Proc. Ntal. Acad. Sci.*, vol. 83, pp. 8390–8394, 1986.

[18] F. Attneave, "Some informational aspects of visual perception," *Psychol. Rev.*, vol. 61, pp. 183–193, 1954.

[19] H. Barlow, "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, W. Rosenblith, Ed. Cambridge, MA: MIT Press, 1961, pp. 217–234.

[20] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[21] A. J. Bell and T. J. Sejnowski, "The 'independent components' of natural scenes are edge filters," *Vision Research*, vol. 37, no. 23, pp. 3327–3338, 1997. [Online]. Available: ftp://ftp.cnl.salk.edu/pub/tony/vis3.ps.Zandhttp://citeseer.nj.nec.com/bell97independent.html

[22] J. J. Atick, "Could information theory provide an ecologial theory of sensory processing?" *Network Computation in Neural Systems*, vol. 3, no. 2, pp. 213–251, 1992.

[23] B. A. Olshausen and D. J. Field, "Vision and the coding of natural images," *American Scientist*, vol. 88, no. 3, pp. 238–245, 2000.

[24] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neuroscience*, vol. 4, no. 8, pp. 819–825, 2001.

[25] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, "Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons," in *Probabilistic Models of the Brain: Perception and Neural Function*, R. Rao, B. Olshausen, and M. Lewicki, Eds. MIT Press, 2001.

[26] E. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, 2001.

[27] J. Malo, F. Ferri, J. Albert, J.Soret, and J. Artigas, "The role of perceptual contrast non-linearities in image transform coding," *Image & Vision Computing*, vol. 18, no. 3, pp. 233–246, 2000.

[28] B. R. Hunt, "Digital image processing," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 693–708, 1975.

[29] J. Karhunen, S. Malariou, and M. Ilmoniemi, "Local linear independent component analysis based on clustering," *International Journal of Neural Systems*, vol. 10, no. 6, pp. 439–451, 2000.

[30] N. Kambhatla and T. Leen, "Dimension reduction by local PCA," *Neural Computation*, no. 9, pp. 1493–1500, 1997.

[31] A. Watson, "Efficiency of a model human image code," *Journal of Optical Society of America A*, vol. 4, no. 12, pp. 2401–2417, 1987.

[32] J. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," *IEEE Transactions on Biomedical Engineering*, vol. 36, pp. 107–114, 1989.

[33] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*. Boston: Kluwer Academic Press, 1992.

[34] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 127–135, 1982.

[35] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Communications of the ACM*, vol. 34, no. 4, pp. 47–58, 1991.

[36] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and new contribution," *Proceedings IEEE*, vol. 83, no. 6, pp. 858–876, 1995.

[37] J. Malo, F. Ferri, J. Albert, and J. Artigas, "Splitting criterion for hierarchical motion estimation based on perceptual coding," *Electronics Letters*, vol. 34, no. 6, pp. 541–543, 1998.

[38] J. Wang and E. Adelson, "Representing moving images with layers," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 625–638, 1994.

[39] L. Torres, D. García, and A. Mates, "A robust motion estimation and segmentation approach to represent moving images with layers," in *IEEE Intl. Conf. Acoust. Speech Sig. Proc.*, Munich, Germany, 1997, pp. 2981–2984.

[40] G. Borgefors, "Distance transformations in digital images," *Computer Vision, Graphics and Image Processing*, vol. 34, pp. 344–371, 1986.

[41] J. Malo, F. Ferri, J. Albert, and J. Soret, "Comparison of perceptually uniform quantization with average error minimization in image transform coding," *Electronics Letters*, vol. 35, no. 13, pp. 1067–1068, 1999.

[42] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Soviet Math. Docl.*, vol. 4, pp. 1035–1038, 1963.

[43] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal of Sci. Comp.*, vol. 14, pp. 1487–1503, 1993.

[44] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 2001.

[45] J. Foley, "Human luminance pattern mechanisms: Masking experiments require a new model," *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1710–1719, 1994.

[46] L. Harvey and V. Doan, "Visual masking at different polar angles in the two-dimensional fourier plane," *Journal of the Optical Society of America A*, vol. 7, no. 1, pp. 116–127, 1990.

[47] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, vol. 58, 1987.

[48] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?" *Vision Research*, vol. 37, pp. 3311–25, 1997.

[49] D. Field, "Wavelets, vision and the statistics of natural scenes," *Phil. Trans. R. Soc. Lond. A*, no. 357, pp. 2527–2542, 1999.

[50] G. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 31–43, 1991.

[51] G. Gómez, G. Camps-Valls, J. Gutiérrez, and J. Malo, "Perceptual adaptive insensitivity for support vector machine image coding," *Aceptado para su publicación en IEEE Transactions on Neural Networks*, 2005.

[52] A. Hyvarinen, J. Hurri, and J. Vayrynen, "Bubbles: a unifying framework for low-level statistival properties of natural image sequences," *Journal of the Optical Society of America A*, vol. 20, no. 7, 2003.

[53] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *To appear in: IEEE Transactions on Image Processing*, 2005.

[54] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.

[55] J. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc.R.Soc.Lond. B*, vol. 265, pp. 359–366, 1998.

[56] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pp. 245–256, 2003.

[57] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[58] W. Press, B. Flannery, S. Teulosky, and W. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge: Cambridge University Press, 1992.

[59] J. Malo, J. Gutierrez, and J. Rovira, "Perturbation analysis of changes in v1 due to context," Gordon Research Conference on Sensory Coding and the Natural Environment, September 2004, oxford.

[60] E. Peli, "Contrast in complex images," *JOSA A*, vol. 7, pp. 2032–2040, 1990.

[61] T. Cover and J. Tomas, *Elements of Information Theory.* New York: John Wiley & Sons, 1991.

[62] J. D. Jackson, *Classical Electrodynamics.* NY: John Wiley and Sons, 1998.

[63] G. Schuster and A. Katsaggelos, "A video compression scheme with optimal bit allocation among segmentation, motion and residual error," *IEEE Transactions on Image Processing*, vol. 6, no. 11, pp. 1487–1502, 1997.

[64] H. Musmann, P. Pirsch, and H. Grallert, "Advances in picture coding," *Proceedings IEEE*, vol. 73, no. 4, pp. 523–548, 1985.

[65] A. Tekalp, *Digital Video Processing.* Upper Saddle River, NJ: Prentice Hall, 1995.

[66] L. Torres and M. Kunt, *Video Coding: A Second Generation Approach.* Boston: Kluwer Academic Publishers, 1996.

[67] T. Sikora, "MPEG digital video-coding standards," *IEEE Signal Processing Magazine*, pp. 82–100, 1997.

[68] J. Jain and A. Jain, "Displacement measurement and its applications in interframe image coding," *IEEE Transcations on Communications*, vol. 29, no. 12, pp. 1799–1808, 1981.

[69] J. Daugman, "Uncertainity relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America*, vol. 2, no. 7, pp. 1160–1169, 1985.

[70] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 58, 1987.

[71] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contours models," *Int. Journal Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[72] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *GVGIP Image Understanding*, vol. 55, no. 1, pp. 14–26, January 1992.

[73] J. Gutiérrez, I. Llorens, F. Ferri, and J. Albert, "Minimización de b-snakes usando algoritmos voraces y distancia chamfer," in *Actas de la VII Conferencia de la AEPIA*, Torremolinos (Málaga), Noviembre 1997, pp. 207–215.

[74] J. Gutiérrez, "Contornos activos: bases teóricas, representación y minimización," Tesina, Universitat de València, Septiembre 1999.

[75] J. Hadamard, *Bull. Univ. Princeton*, vol. 13, no. 49, 1902.

[76] ——, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, 1923.

[77] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 26, pp. 314–319, 1985.

[78] M. Bertero, "Regularization methods for linear inverse problems," in *Lecture Notes in Mathematics.* Berlin: Springer, 1986, no. 1225, pp. 53–112.

[79] J. Biemond, R. L. Jagendijk, and R. M. Merserau, "Iterative methods for image deblurring," *Proceedings of the IEEE*, vol. 78, pp. 856–883, 1990.

[80] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Sig. Proc. Mag.*, vol. 14, pp. 24–41, 1997.

[81] A. Björck and L. Eldén, "Methods in numerial algebra for ill-posed problems," in *International Symposium on Ill-Posed Problems: Theory and Practice*, 1979.

[82] J. M. Varah, "A practical examination of some numerical methods for linear discrete ill-posed problems," *SIAM Review*, vol. 21, no. 1, pp. 100–112, 1979.

[83] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems.* SIAM, 1998.

[84] M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869–889, 1988.

[85] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* NJ: Prentice Hall, 1977.

[86] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[87] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 20, no. 6, pp. 1964–1977, 1999.

[88] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the l-curve," *SIAM Rev.*, vol. 34, pp. 561–580, 1992.

[89] R. Molina, J. Núñez, F. J. Cortijo, and J. Mateos, "Image restoration in astronomy," *IEEE Sig. Proc. Mag.*, pp. 11–29, March 2001.

[90] D. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[91] J. Lim, *Two-Dimensional Signal and Image Processing.* Prentice Hall, 1990.

[92] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[93] J. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Pat. Anal. Mach. Intell.*, vol. 2, pp. 165–168, 1980.

APÉNDICE

A. TRABAJO 1

V1 non-linear properties emerge from local linear ICA

Jesús Malo, **Juan Gutiérrez** y José Rovira.

En proceso de revisión en *Network: Computation in Neural Systems*

V1 non-linear properties emerge from local linear ICA[1]

J. Malo[2] , J. Gutiérrez⋆, J. Rovira

Dept. d'Òptica, Facultat de Física, Universitat de València
(⋆) Dept. d'Informàtica, Escola Superior d'Enginyeries, Universitat de València
Dr. Moliner 50, 46100 Burjassot, València, (Spain)

It has been argued that the aim of the non-linearities in different visual and auditory mechanisms may be removing the relations between the coefficients of the signal after global linear ICA-like stages (Schwartz and Simoncelli, 2001). Specifically, the authors showed that the above strategy, i.e. removing the bow-tie like dependencies between wavelet coefficients, was appropriate to obtain the parameters of a particular non-linearity that reproduces masking effects.
In this work we present a different result that supports the same efficient encoding hypothesis: instead of assuming a particular functional form for the non-linearity, we show that using a local-to-global non-linear ICA masking-like behavior is reproduced as well. This result is an additional indication that Barlow's efficient encoding hypothesis may explain not only the shape of receptive fields of V1 sensors but also their non-linear behavior.

Keywords: Human Vision, Natural Image Statistics, ICA, Wavelets, Non-linear Systems, Incremental Thresholds, Gain Control, Divisive Normalization.

[1]This work was partially supported by the projects CICYT TIC2003-01504, Grupos04-08 (GV), and by the program DN 8644.

[2]Corresponding author: Jesús Malo. Dept. d'Òptica, Facultat de Física, Universitat de València. Dr. Moliner 50, 46100 Burjassot, València, (Spain). Tel: +34 963544717, Fax: +34 963544715, e-mail: Jesus.Malo@uv.es

V1 non-linear properties emerge from local linear ICA

1 Introduction

The different specific formulations of Barlow's idea about the match between perception and the statistics of natural signals (Barlow, 1961; Barlow, 2001) have explained a variety of perception facts (Simoncelli and Olshausen, 2001; Simoncelli, 2003). In its original formulation (Barlow, 1961), Barlow suggested that the goal of low-level perception mechanisms is obtaining a set of statistically independent responses. Recently (Barlow, 2001), he generalized the relevance of statistics in perception to include more abstract (bayesian) representations of the information about the environment. Nowadays there is a fruitful debate about the generality of the original efficient encoding hypothesis (Barlow, 2001; Simoncelli, 2003).

However, despite the (eventually) restricted applicability of the original efficient encoding formulation, it had remarkable success in explaining the particular shape of linear V1 receptive fields (Olshausen and Field, 1996). Moreover, in (Schwartz and Simoncelli, 2001) the same idea has been used to obtain the parameters of a particular functional form of the non-linearities found in V1: the divisive normalization (Heeger, 1992).

In this work we show that applying a particular (local-to-global) non-linear ICA, that assumes no specific functional form for the non-linearity, the masking-like non-linearities are reproduced as well. Then, the statistical role of the non-linearities would be adapting the global representation of the previous linear stage (global linear ICA) to the local features of the signal around each observed (masking) stimulus. This result is an additional indication that the original efficient encoding formulation of Barlow ideas may explain not only the linear behavior of V1, but also its non-linearities. The difference of the result presented here with regard to previous implementations of the efficient encoding idea (Schwartz and Simoncelli, 2001) is that no explicit functional form is assumed for the non-linearity, thus revealing more clearly that it emerges from the particular PDF of natural images.

The structure of the paper is as follows. In Section 2, we review the relations between local linear ICA representations and the more general non-linear ICA representations (Hyvärinen et al., 2001; Lin, 1999; Karhunen et al., 2000; Jutten and Karhunen, 2003), the particular local-to-global non-linear ICA used in this paper is derived from these relations. In Section 3, we apply the above

concepts to the current V1 model. In this way, we derive an integral expression for the non-linear part of the current V1 model. This expression assumes no particular functional form for the non-linearity. In Section 4, we show that when using the above expression with natural images, different masking behaviors are reproduced. Finally, in Section 5 we present the conclusions of the work.

2 Global linear ICA, local linear ICA and non-linear ICA

The basic data model used in (global) linear ICA assumes that the n-dimensional samples of the signal, the column vectors $\mathbf{x}$, come from a mixture of m independent sources, column vectors $\mathbf{r}$, given by the (global) mixing matrix, $\mathbf{A_g}$ (Hyvärinen et al., 2001):

$$\mathbf{x} = \mathbf{A_g} \cdot \mathbf{r} = \sum_{j=1}^{m} \mathbf{A_g}_j r_j \tag{1}$$

The column vectors in the mixing matrix, $\mathbf{A_g}_j$, $j = 1, \cdots, m$, are the basis vectors of ICA. Assuming that the mixing matrix of the model is unknown, the independent components are found by determining a linear inverse mapping (separating matrix), $\mathbf{W_g}$, so that the m-vector

$$\mathbf{r} = \mathbf{W_g} \cdot \mathbf{x} \tag{2}$$

is an estimate of the independent component vector. A number of different algorithms have been proposed to compute $\mathbf{W_g}$ (Hyvärinen et al., 2001). For simplicity the number of independent components, m, is usually assumed to be equal to the dimension of the data, so the mixing matrix is simply the inverse of the separating matrix, $\mathbf{A_g} = \mathbf{W_g}^{-1}$.

In spite of its usefulness, the basic data model assumed in eq. 1 is often too simple to describe real-world data. First, the standard model is *linear*, while generally we need a suitable *non-linear* model to adequately represent the data. Second, the ICA model in eq. 1 is *global*, i.e. it tries to describe all the data using the same global features (ICA basis vectors, $\mathbf{A_g}_j$). However, natural data often have different *local* features in different parts of its global PDF. This is why non-linear ICA methods (Hyvärinen et al., 2001; Jutten and Karhunen, 2003) and local linear ICA methods (Hyvärinen et al., 2001; Karhunen et al., 2000) have been developed.

General non-linear ICA techniques (Hyvärinen et al., 2001; Jutten and Karhunen, 2003) assume that the data vectors, $\mathbf{x}$, depend non-linearly on some statistically independent components, $\mathbf{r}$,

$$\mathbf{x} = F(\mathbf{r}) \tag{3}$$

and then, again, the problem is obtaining the inverse mapping,

$$\mathbf{r} = F^{-1}(\mathbf{x}) \tag{4}$$

that estimates the independent components. The problems with this generic formulation are, (1) it is computationally expensive, and (2) it is highly non-unique (Jutten and Karhunen, 2003): i.e. multiple inverse mappings may be obtained that fulfill the condition of giving rise to vectors $\mathbf{r}$ with independent components.

Local linear methods (Hyvärinen et al., 2001; Karhunen et al., 2000; Kambhatla and Leen, 1997) start from finding clusters in the global PDF, and then, a different *local* linear model is applied to the data in each cluster. The clustering part of these techniques is the way to break the (difficult) global non-linear problem into a set of (easier) local subproblems where the linear approach is valid enough.

However, a different approach is possible to relate the global and the local description of the data. As stated in (Lin, 1999; Karhunen et al., 2000), if the non-linear mapping in eq. 3 is smooth, then, a local (differential) approach can be taken, considering the Taylor expansion of F up to first order. In this case, the mixing non-linearity of the model at the point $\mathbf{r}'$ (or equivalently, at the point $\mathbf{x}' = F(\mathbf{r}')$), could be locally approximated by a *local linear* mixing matrix, $\mathbf{A}(\mathbf{x}')$,

$$d\mathbf{x}' = \nabla F(\mathbf{r}') \cdot d\mathbf{r}' = \mathbf{A}(\mathbf{x}') \cdot d\mathbf{r}' \tag{5}$$

Accordingly, the inverse (separating) mapping, would be locally approximated by,

$$d\mathbf{r}' = \nabla F^{-1}(\mathbf{x}') \cdot d\mathbf{x}' = \mathbf{W}(\mathbf{x}') \cdot d\mathbf{x}' \tag{6}$$

In this case, the local features $\mathbf{A}(\mathbf{x}')$ could be derived from the (local) $\mathbf{W}(\mathbf{x}')$, applying to local neighborhoods the same techniques that are used in the global case, as done in the linear local approaches that use pre-clustering (Kambhatla and Leen, 1997; Karhunen et al., 2000).

In this paper we consider the latter (differential) way to relate the local linear analysis of the signal to the global non-linear analysis. In this case, the global representation can be obtained by integrating the local increments,

$$\mathbf{r} = \mathbf{r_0} + \int_{\mathbf{x}_0}^{\mathbf{x}} \nabla F^{-1}(\mathbf{x}') \cdot d\mathbf{x}' = \mathbf{r_0} + \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{W}(\mathbf{x}') \cdot d\mathbf{x}' \tag{7}$$

where the local separating matrix at each point is computed using standard linear ICA techniques (e.g. the symmetric fixed-point algorithm (Hyvärinen et al., 2001; Hyvärinen, 1999)) in a neighborhood of the point $\mathbf{x}'$. While in (Karhunen et al., 2000), the local expansions are defined by the

set of data vectors belonging to each cluster, in our case, they are defined by the way we choose the neighborhood. In particular, our local neighborhoods are formed by the K closer neighbors in Euclidean sense.

The particular local-to-global non-linear ICA representation given by eq. 7 is useful only if (1) the final components of $\mathbf{r}$ are more independent than the global ICA components, and (2) the result of the integral is independent of the integration path, so no particular path has to be specified. In the appendices A and B, we show that this is the case in the particular problem of interest: the PDF of natural images.

3 V1 non-linearities and local-to-global non-linear ICA

The current model of V1 sensors involves a *linear stage* and a *non-linear stage* (Schwartz and Simoncelli, 2001; Watson and Solomon, 1997).

First, the input images, $\mathbf{x}$, are analyzed by a set of *unit-norm* linear sensors:

$$\mathbf{c} = \mathbf{W_{gu}} \cdot \mathbf{x} \tag{8}$$

Then, a non-trivial non-linearity is applied to the outputs of the linear stage:

$$\mathbf{r} = R(\mathbf{c}) \tag{9}$$

The whole set of transforms (with the corresponding inverses) can be summarized as follows:

$$\mathbf{x} \underset{\mathbf{A_{gu}}}{\overset{\mathbf{W_{gu}}}{\rightleftarrows}} \mathbf{c} \underset{R^{-1}}{\overset{R}{\rightleftarrows}} \mathbf{r} \tag{10}$$

The challenge of statistically based vision models is to derive the experimental behavior of these transforms from the efficient encoding hypothesis together with the statistical properties of natural images. As stated in the introduction, the receptive fields of the sensors of the first linear stage resemble the global linear ICA basis functions of natural images (Olshausen and Field, 1996). Accordingly, we used the notation $\mathbf{W_{gu}}$ and $\mathbf{A_{gu}}$ for these (global) transforms. However, we added the subscript $\mathbf{u}$ to stress the difference with regard to standard (global) ICA techniques that give rise to basis functions of different norm (Hyvärinen, 1999). In this case,

$$\mathbf{W_{gu}} = D_g^{\frac{1}{2}} \cdot \mathbf{W_g} \tag{11}$$

$$\mathbf{A_{gu}} = \mathbf{A_g} \cdot D_g^{-\frac{1}{2}} \tag{12}$$

where the diagonal matrix $D_g^{\frac{1}{2}}$ contains the norms of the (global) vectors in $\mathbf{A_g}$. Note that the application of this diagonal matrix doesn't modify the independence properties of the original global transform.

As stated in Section 2, in general, global linear ICA techniques do not give rise to truly independent coefficients. Assuming the original formulation of Barlow ideas, the role of the non-linear transform, R, would be removing the remaining relations between the coefficients of the vector $\mathbf{c}$. Following this hypothesis, Schwartz and Simoncelli (Schwartz and Simoncelli, 2001) fitted the parameters of a particular model for the non-linearity (the divisive normalization (Heeger, 1992)) to obtain independent responses in the vector $\mathbf{r}$. By doing so, they reproduced the experimental masking-like non-linearities.

In this paper we follow the same general idea (obtain the non-linearity, R, by looking for independent components in $\mathbf{r}$), but assuming no particular functional form in R and using the local-to-global ICA described in Section 2.

To this end, we take a differential approach to locally approximate the non-linear response, R, by its Jacobian[3]. Then, given a masking stimulus, $\mathbf{x}'$, the increment in the response of the sensors given a variation of the stimulus in the pixel domain is:

$$d\mathbf{r}' = \nabla R(\mathbf{c}') \cdot d\mathbf{c}' = \nabla R(\mathbf{W_{gu}} \cdot \mathbf{x}') \cdot \mathbf{W_{gu}} \cdot d\mathbf{x}' \tag{13}$$

so the (global) response to any stimulus, $\mathbf{x}$, is just:

$$\mathbf{r} = \mathbf{r_0} + \int_{x_0}^{x} \nabla R(\mathbf{W_{gu}} \cdot \mathbf{x}') \cdot \mathbf{W_{gu}} \cdot d\mathbf{x}' \tag{14}$$

Assuming that the role of the cascade of linear and non-linear transforms in eq. 10 is obtaining independent components in the coefficients of the response vector $\mathbf{r}$, the local behavior of the non-linearity could be obtained identifying eqs. 6 and 13. From this identification, we get,

$$\nabla R(\mathbf{W_{gu}} \cdot \mathbf{x}') = \mathbf{W}(\mathbf{x}') \cdot \mathbf{A_{gu}} \tag{15}$$

that applied into eq. 14, gives rise to eq. 7. Eq. 15 states that the role of the V1 non-linearities is representing each input image (each masking stimulus $\mathbf{x}'$) into the *local linear* representation that captures the *local features* of the natural images around it.

[3]Preliminar results of the differential approach used in this paper were presented in the GRC meeting (Malo et al., 2004).

4 Perception facts, implementation details and simulation results

In this section we apply eq. 15 and its consequence (eq. 7) to a set of natural images. The results obtained show that the response of the sensors constructed in this way reproduces the basic behavior of V1 sensors as revealed by threshold and suprathreshold psychophysics (Daugman, 1980; Watson, 1983; Harvey and Doan, 1990; Legge and Foley, 1980; Legge, 1981; Foley, 1994; Watson and Solomon, 1997).

4.1 Perception facts to be (statistically) reproduced

- **Linear result**:
  - *Set of linear frequency analyzers*: The linear receptive fields of the V1 sensors are narrow band-pass functions that analyze the frequency content of the images in a log-polar wavelet-like way (Daugman, 1980; Watson, 1983; Harvey and Doan, 1990).
- **Non-linear results**:
  - *Auto-masking*: Contrast incremental thresholds of sensors tuned to specific frequency bands increase with the amplitude (or contrast) of the stimulus (Legge and Foley, 1980; Legge, 1981). This is equivalent to say that (1) the sensitivity of the mechanisms decreases with the contrast of the optimal stimulus, or (2) the response of the mechanisms when excited with their optimal stimulus is a saturating non-linearity.
  - *Cross-masking I, General behavior*: The sensitivity of a particular sensor to its optimal stimulus changes (decreases) when this stimulus is superimposed on a high contrast background of different frequency content (cross-masking) (Foley, 1994; Watson and Solomon, 1997). This effect increases with the contrast of the background.
  - *Cross-masking II, Fine details, relative influence between frequencies*: Moreover, the above effect (or perceptual interaction between frequencies) decreases as the distance between the frequencies (in modulus or orientation) increases (Watson and Solomon, 1997).

In this work, our aim is limited to the non-linear effects because the linear effects were previously explained by others using a number of concurrent statistical arguments (e.g., Field (Field, 1987),

Olshausen (Olshausen and Field, 1996) or Atick (Atick, 1992)). For instance, as stated above, the reproduction of the linear receptive fields is not a merit of the proposed procedure, but a result of the *assumed* initial global ICA stage (Olshausen and Field, 1996).

4.2 Implementation details

In order to reduce the complexity of the simulations and to make the visualization easier, we restrict ourselves to low-dimensional (6D) examples. In all cases the samples in the pixel domain were taken from a 1D subspace (a line) of the spatial domain. Of course, this prevent us to analyze orientation effects because to do so, you at least need 4 (2×2) or 9 (3×3) two-dimensionally taken samples. The problem of doing so is that in those (low dimensionality) two-dimensionally taken examples, the number of different moduli for the spatial frequency is very small, so the cross-masking results will be more difficult to analyze. This is why we restrict ourselves here to one-dimensionally taken examples.

Of course all these problems should disappear using substantially bigger image blocks (say 8×8 or 12×12). We preferred not to do that because of computational expenses. However, it has to be stressed that the computational problem we face here is not as fundamental as the classical course-of-dimensionality problem you find in multidimensional PDF estimation. Here the problem is not that serious and alternative solutions (local-PCA based) are available (Kambhatla and Leen, 1997).

Specifically, the problem here is the numerical (Runge-Kutta) integration of eq. 7. There the bottleneck is the speed of the algorithm to compute $\mathbf{W}$. We are using the Matlab implementation of FastICA (Hyvärinen, 1999), which converges slowly for high-dimensional vectors. However, this (small) computational problem could be solved by using local-PCA methods (Kambhatla and Leen, 1997) in the proposed local-to-global framework. Preliminary 6D experiments with local PCA show that the results slightly differ from the local ICA results. Therefore, for consistency with the first (global linear) part, we also use (local linear) ICA to reproduce the non-linearity instead of choosing the (much faster!) local PCA.

However, note that the above considerations do not restrict the generality of the proposed approach. The extension of the procedure to higher dimensions may introduce more noise in the local linear functions but this is not a fundamental problem either because it can be easily solved by considering more images in the training set. The aim of this work is reporting the general

local-to-global framework and its encouraging results for statistically based vision: the particular lower bounds in the amount of data to be considered in order to make the method efficient and robust in highly multidimensional situations are a matter for future research.

The examples shown in the sections below were carried out using $2 \cdot 10^6$, 6×1, image samples randomly taken from a well known natural image database (van Hateren and van der Schaaf, 1998). The original images of the database were scaled to lie in the range $[0, 255]$. The local neighborhoods to compute the local ICAs were formed by the $5 \cdot 10^4$ closer samples around each point.

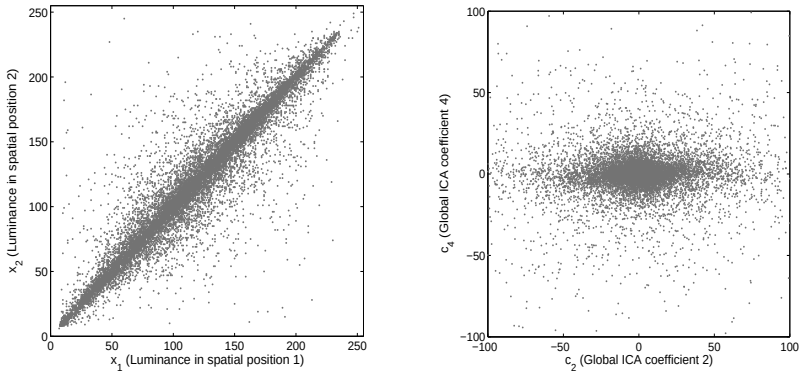
4.3 Simulation results

4.3.1 Data and global transform

First we show the kind of raw data we used in this set of experiments and the basis functions of the global linear transform, $\mathbf{W_{gu}}$ and $\mathbf{A_{gu}}$, obtained using ICA as proposed by Olshausen and Field (Olshausen and Field, 1996).

Figure 1.a shows a projection (first two components) of a subset of the data in the spatial domain representation. It is clear that $\mathbf{x_i}$ coefficients are highly redundant. Figure 2 shows the different basis functions (the columns of $\mathbf{A_{gu}}$) in the spatial domain, and their frequency content. These functions were computed using the standard symmetric fixed-point ICA algorithm (Hyvärinen et al., 2001; Hyvärinen, 1999).

This set of linear sensors would correspond to the set of V1 linear frequency analyzers (the



a) b)

Figure 1: Samples of the natural data in, a) the spatial domain, and b) the global linear ICA domain.

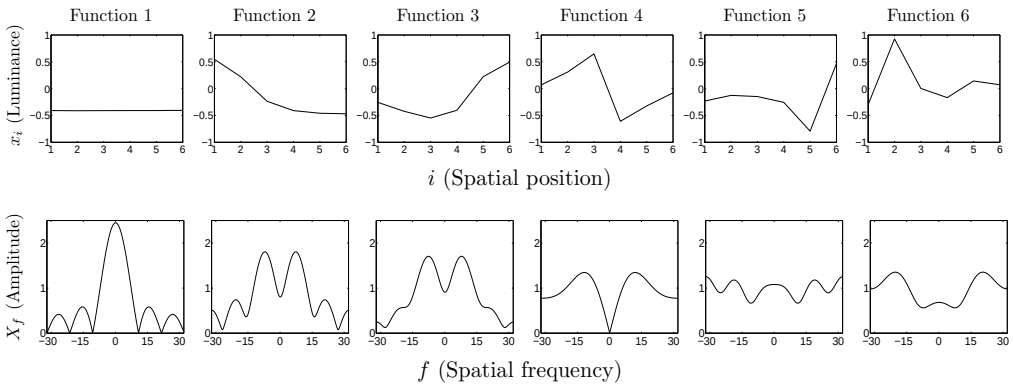


Figure 2: Basis functions of the global ICA transform in the spatial domain, x_i, and their frequency content, X_f. The functions in $\mathbf{A_{gu}}$ were sorted according to the variance of the coefficients in the transformed domain, which is equivalent to sort them according to the norm of the basis functions in $\mathbf{A_g}$.

experimental linear result) quoted above (Olshausen and Field, 1996). Figure 1.b clearly shows that the responses of the above linear sensors, $\mathbf{c_i}$, are highly more independent than the original coefficients $\mathbf{x_i}$.

4.3.2 Reproducing auto-masking

The non-linear properties of a sensor are (psychophysically) revealed by measuring contrast incremental thresholds of patterns that ideally isolate the response of the sensor. In this kind of experiments, the observer has to detect the presence of an incremental stimulus, $\Delta\mathbf{x}$, shown on top of a background (or masking stimulus) $\mathbf{x_0}$.

In order to analyze the response of a particular sensor, i, optimal incremental stimuli should be ideally designed to isolate its response. Therefore, they should lie in the direction that modify a single coefficient of the response vector. In these experiments, the contrast of the incremental stimulus in the optimal direction is increased until the distortion is just noticeable. The contrast of this just noticeable optimal incremental stimulus, $\Delta\mathbf{x}^{(i)}(\mathbf{x_0})$, is referred to as contrast incremental threshold of the sensor i adapted to the masking stimulus $\mathbf{x_0}$. Mathematically, $\Delta\mathbf{x}^{(i)}(\mathbf{x_0})$, is the stimulus that modifies the response by $\Delta\mathbf{r}(\Delta\mathbf{x}^{(i)}) = \tau \cdot \delta^{(i)}$, i.e. $\Delta\mathbf{r}_j(\Delta\mathbf{x}^{(i)}) = \tau \cdot \delta_{j-i}$.

Auto-masking experiments analyze the response of a sensor i (i.e., explore the incremental thresholds in the direction $\Delta\mathbf{x}^{(i)}$), using masking stimuli, $\mathbf{x_0}$, that only excite the same sensor, i.

In order to reproduce the experimental non-linear effects first we have to obtain the expression of the optimal incremental stimulus that isolate the response of a single non-linear sensor. Assuming the definition of the optimal distortion in the direction i, and using eqs. 13 and 15, we have:

$$\Delta \mathbf{x}^{(i)}(\mathbf{x_0}) = \mathbf{A_{gu}} \cdot \nabla R^{-1}(c_0) \cdot \tau \cdot \delta^{(i)} = \tau \cdot \mathbf{A}(\mathbf{x_0}) \cdot \delta^{(i)} \tag{16}$$

As expected from eq. 15, given a certain masking stimulus, $\mathbf{x_0}$, the non-linear sensor, i, is optimally excited by the column, i, of the local-linear ICA matrix, $\mathbf{A}(\mathbf{x_0})$.

In order to establish a consistent definition of the non-linear sensors, some fixed criterion has to be applied to sort (or enumerate) the basis functions of the different local representations. In our case, first we computed the local basis functions around the origin (see fig. 3) and we sorted them according to their norm (as in the example of fig. 2). Then, for any other neighborhood, the local basis were compared to the basis around the origin using the inner product, and they were named (sorted) accordingly.

The above definition of optimal incremental stimuli is convenient to simulate the auto-masking results: note that starting from the average image (zero contrast, mean luminance background), if one progressively changes it according to the local optimal incremental stimulus for a particular sensor, one obtains constant increments in this single sensor, thus reproducing the ideal conditions of the auto-masking experiments. Figure 4 shows an illustrative example of the directions (stimuli) that isolate the response of the 2nd and 4th non-linear sensors starting from the zero-contrast

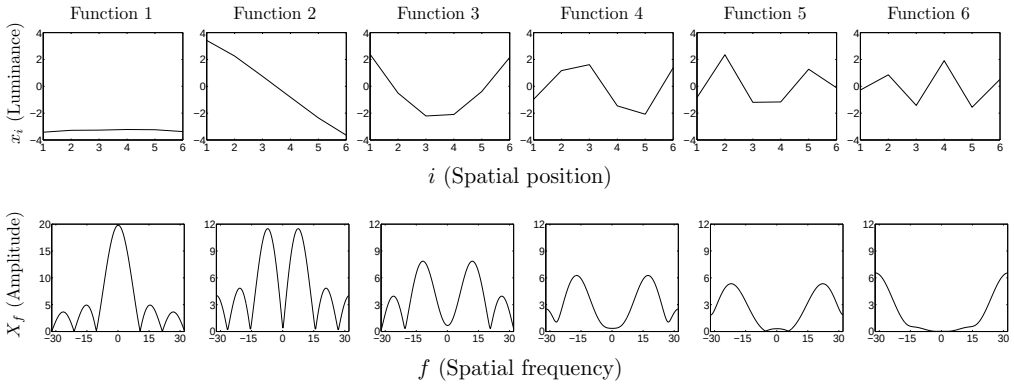


Figure 3: Basis functions of the local ICA representation (around the average stimulus) in the spatial domain, x_i, and their frequency content, X_f.

stimulus (the average image).

First, note how these directions do not exactly coincide with the 2nd and 4th axes of the linear representation. Second, note how the length of the basis functions (and hence the contrast incremental thresholds) increase as one moves away from the origin (as the contrast of the stimulus increases). This is clearer in fig. 5, that shows different masking stimuli isolating the 4th sensor and the corresponding incremental stimuli in the spatial domain.

In each case, the new masking stimulus is computed modifying the previous one (increasing the

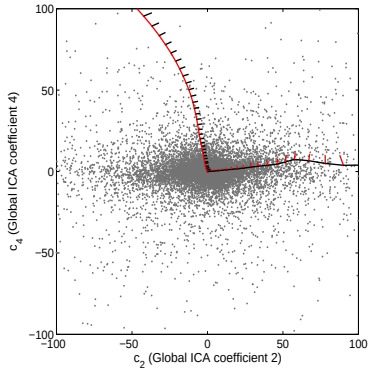
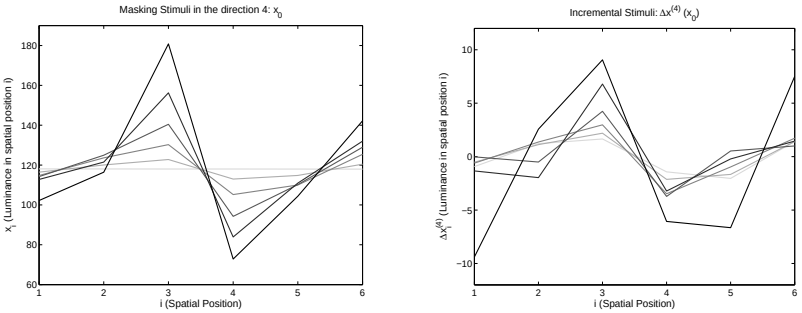


Figure 4: Directions that isolate the response of sensors 2 (black) and 4 (red). In each point of these optimal stimuli paths, the 2nd and 4th local basis functions have been plotted in black and red respectively.



a) b)

Figure 5: a) Masking stimuli of increasing contrast (gray-to-black) that only excite the 4th sensor. b) Optimal incremental stimuli $\Delta\mathbf{x}^{(4)}(\mathbf{x_0})$ for the masking stimuli of figure a). Note that as the contrast of the masking stimulus increases, the contrast of the optimal incremental stimulus (the contrast incremental threshold), also increases.

contrast) in the direction that isolates the response of the selected mechanism. The increment in the response when modifying the stimuli from one point to the next is $\tau \cdot \delta^{(i)}$. Then, the Michaelson contrast, ΔC, of the computed incremental stimulus is the necessary contrast increment to produce a threshold variation, τ, in the response. We use the (variable) increments in contrast, ΔC, and the (constant) increments in the response, τ, to derive the non-linear response versus the Michaelson contrast. Figure 6 shows the response of the 4th sensor in auto-masking conditions computed following the above procedure. Note that the contrast incremental threshold (the contrast of the incremental stimuli, ΔC) increase with the contrast of the masking stimulus. The same kind of saturating non-linear response is obtained for all the AC sensors when excited by their optimal stimuli (see the auto-masking –solid– curves in figure 8).

These results suggest that the proposed local-to-global ICA sensors reproduce the basic V1 non-linear behavior in auto-masking conditions. In the proposed local-to-global ICA the response is a saturating non-linearity (the sensitivity of the sensors decrease with contrast) because the radius of the neighborhood increases to compensate the decrease of the PDF for high contrast images. The point-wise non-linear result of auto-masking is similar to an optimal distribution of just noticeable distortions (discrete perceptions or quantization levels) for minimum MSE, which gives rise to smaller incremental thresholds in high probability regions and viceversa (Lloyd, 1982; Watson, 1987; Gersho and Gray, 1992; Malo et al., 2000).

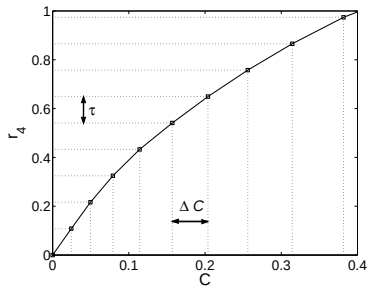


Figure 6: Non-linear response of the 4th sensor in auto-masking conditions. At different Michaelson contrast levels, C, in the optimal direction different contrast incremental thresholds, ΔC, are needed to generate the same increment τ.

4.3.3 Reproducing cross-masking

In this case, we analyze the contrast of the optimal incremental stimuli in the direction i shown on top of stimuli that excite a different sensor $k \neq i$. To do so, we start by choosing some masking stimulus of a certain contrast in one of the directions revealed by the auto-masking analysis (direction k). Then, we modify the selected masking stimulus in the direction of other sensor (sensor i). Again, the response curve in this situation is obtained from the (variable) contrast of the incremental stimuli, ΔC, and the (constant) increments τ.

Figure 7 shows an example of the above procedure: different masking stimuli (of different contrast) exciting the 2nd sensor (black dots) were selected as starting points for contrast incremental thresholds computations in the 4th direction (red lines). In this case, we chose the starting contrast of the masking stimuli to be 0 (auto-masking), 0.1 (low contrast), 0.2 (medium contrast), and 0.4 (high contrast).

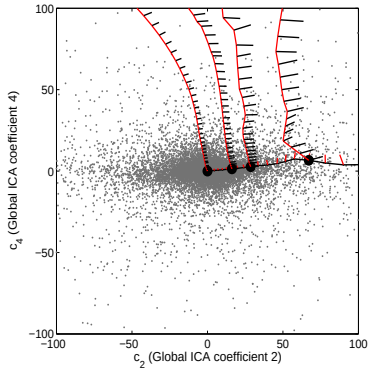


Figure 7: Directions that isolate the response of the 4th sensor (red lines) for different masking stimuli exciting the 2nd sensor (black dots).

Note that as the contrast of the masking stimulus (in the 2nd direction) increases, the contrast incremental thresholds of the 4th sensor also increase (the length of the 4th basis function is bigger), revealing cross-masking between these two directions. The effect of the non-linearity is mapping the non-uniform grid (in the $\mathbf{c}$ domain) given by the directions that isolate the response of particular sensors into a uniform cartesian grid (in the $\mathbf{r}$ domain).

Figure 8 shows the responses of low, medium and high frequency sensors (2nd, 4th and 6th) in auto-masking and cross-masking conditions, using different contrasts (namely 0.1, 0.2 and 0.4)

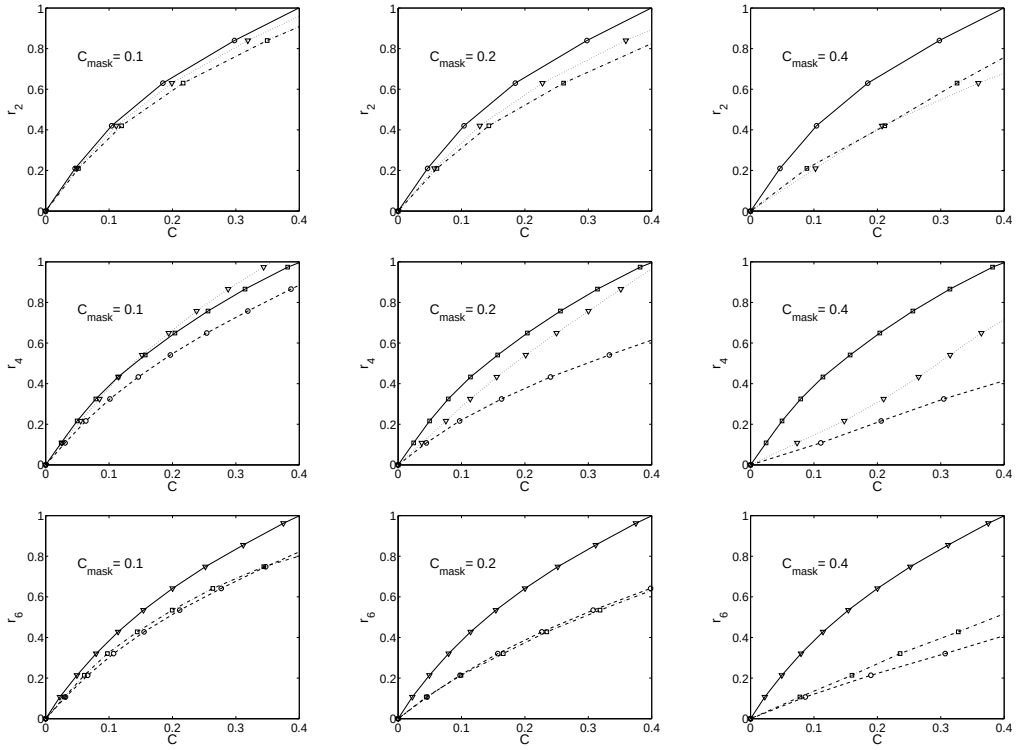


Figure 8: Non-linear response of the low-frequency (first row), medium-frequency (second row) and high-frequency (third row) mechanisms (2nd, 4th and 6th sensors), when using different masking stimuli with different contrast: low contrast (first column), medium contrast (second column) and high contrast (third column). In each case, the solid line represent the response in auto-masking conditions and it is plotted for reference purposes. The other curves represent the response of the considered mechanism when a masking stimulus of low (dashed circles), medium (dash-dot squares) or high frequency (dotted triangles) is present.

for the masking stimuli.

These results show that auto-masking and general cross-masking behavior emerge from natural data using the proposed algorithm: specifically, the response of all sensors when excited by their optimal stimulus on a uniform background (auto-masking) is a saturating non-linearity. Then, if these stimuli are shown on a background of different frequency (exciting some other different sensor) the sensitivity is reduced and this masking effect increases with the contrast of the background stimulus. These general results are quite independent of the selected set of natural images.

However, the fine details of cross-masking, namely the relative interaction between frequen-

cies (Watson and Solomon, 1997), is not clearly obtained in these 6D experiments: for some sets of images the interactions do decrease as the (frequency) distance between masking stimulus and sensor increase, but this effect strongly depends on the selected set of images or the considered sensors. This fact may be due to the low dimensionality of the samples considered in the experiments: this gives rise to basis functions with a strong overlapping in the spatial and the frequency domain (see fig. 3) in such a way that the effective distance between the all linear sensors is small so the relative interaction between them is quite similar.

5 Conclusions and further work

In this work we derived a non-parametric expression for the non-linear V1 behavior based in a local-to-global non-linear ICA. The proposed local framework allows to obtain the global non-linearities of V1 directly from the statistical properties of natural images without an explicit expression (such as the divisive normalization). In the proposed scheme, the aim of the local behavior of the non-linear sensors would be capturing the local features of natural images (local ICA axes) around each masking stimulus.

The proposed set of sensors reproduce auto-masking and the general trends of cross-masking of V1 sensors. These results (consistent with those presented by Schwarz and Simoncelli (Schwartz and Simoncelli, 2001)) are an alternative indication that Barlow's efficient encoding hypothesis may explain not only the shape of receptive fields of V1 sensors but also their non-linear behavior.

Further work should extend the proposed procedure to higher dimensions in order to (1) reproduce orientation effects, and (2) address the fine details of cross-masking (relative influence between linear sensors). In this case, the non-linar sensors should use local-PCA instead of local ICA in order make the method computationally affordable.

A Efficiency of local-to-global non-linear ICA for natural images

In this appendix we show how the proposed local linear ICA increases the independence of the coefficients obtained by a standard global linear ICA technique (Hyvärinen, 1999). This result is consistent with similar results obtained for the *particular* non-linearity based on the divisive normalization model (Schwartz and Simoncelli, 2001; Malo et al., 2005).

We compare the independence of coefficients in the standard global linear ICA domain and in the proposed local linear ICA domain using two measures, (1) bow-tie relations in the conditional

PDFs (Schwartz and Simoncelli, 2001; Hyvärinen et al., 2003), and (2) relative mutual information (Cover and Tomas, 1991; Malo et al., 2005). The results for the spatial domain representation are also included for reference purposes. In these experiments, we used 10^4 6D image samples randomly chosen from the $2 \cdot 10^6$ data set used above. These vectors were linearly transformed using the global ICA separating matrix computed from the whole data set. Then, the non-linear response for these points was computed using eq.7, integrating from the zero contrast stimulus in a straight line integration path.

Statistical dependence between the coefficients in a representation can be qualitatively revealed by analyzing the conditional histograms of pairs of samples of the signal in that domain (Schwartz and Simoncelli, 2001; Hyvärinen et al., 2003). Figure 9 show examples of these conditional histograms for two pairs of coefficients in the spatial, the global linear ICA and in the proposed domain.

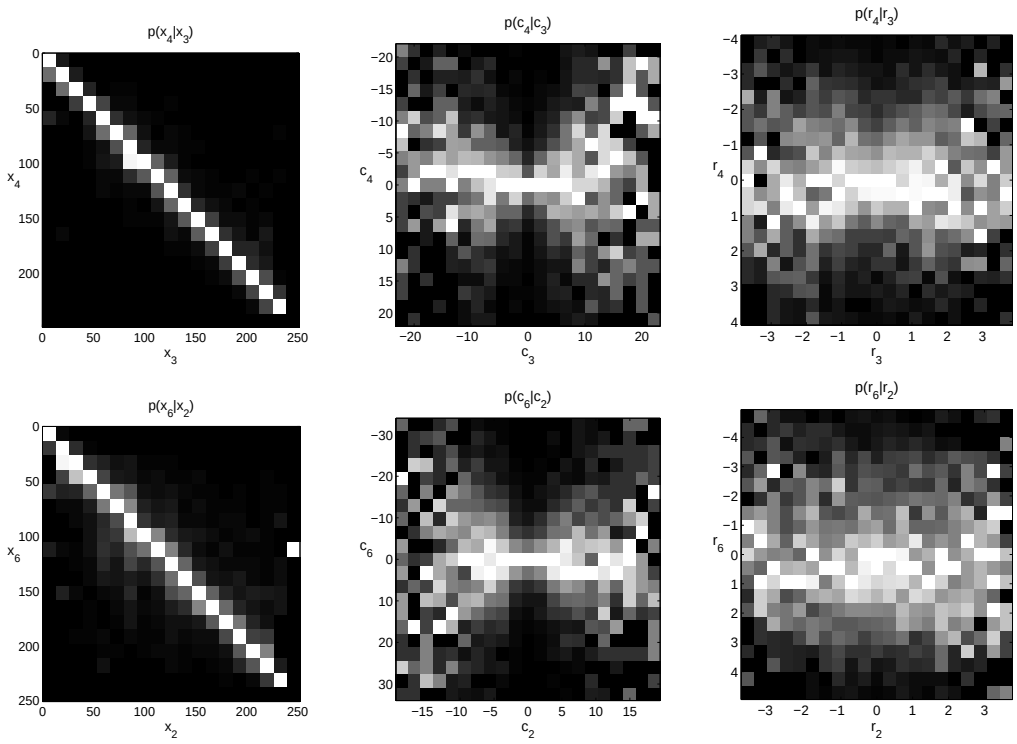


Figure 9: Examples of conditional probabilities of two pairs of coefficients in the three considered domains: Spatial domain (first column), Global linear ICA (second column), local linear ICA (third column).

In the spatial domain case, these results indicate a strong correlation between neighboring luminance values: the average value of x_i changes with the value of its neighbor x_j. In the global linear ICA case, the 2nd order correlation has been removed, note that the average in c_i is independent of the given value of c_j. However, the variance (or energy) of global linear ICA coefficients, c_i, still depend on the value of the neighbors c_j (Schwartz and Simoncelli, 2001; Hyvärinen et al., 2003). In the proposed representation, these bow-tie dependencies are almost removed, thus revealing a weaker statistical dependency between coefficients.

These qualitative results are consistent with quantitative measures based in mutual information. The mutual information of a set of variables, $c_1, \cdots, c_n$, is defined as the Kullback-Leibler distance between their joint PDF and the product of their marginals, and it can be computed from the marginal entropies, $H(c_i)$, and the joint entropy, $H(c_1, \cdots, c_n)$, of the variables (Cover and Tomas, 1991):

$$I(c_1, \cdots, c_n) = \sum_{i=1}^{n} H(c_i) - H(c_1, \cdots, c_n). \tag{17}$$

Mutual information is very useful to know how much information (in bits) can the neighbors tell us about a particular coefficient, and hence, how strong the statistical interactions are in a particular domain. However, some particular value of mutual information can have different interpretations in terms of statistical dependency as a function of the entropy of the considered variables. This is because mutual information is affected by the shape of the marginal PDF (or, equivalently, the marginal entropy) of the coefficients of the representation. A straightforward use of mutual information, I, in order to assess the statistical dependency between variables in different representations will be biased in favor of representations with highly non-uniform marginal PDFs, i.e., representations where the coefficients have small entropy.

Therefore, it is better to define a relative measure that takes the different marginal entropies into account (Malo et al., 2005). We compute the relative mutual information as,

$$I_r(c_1, \cdots, c_n) = \frac{\frac{1}{(n-1)} I(c_1, \cdots, c_n)}{\frac{1}{n} \sum_{i=1}^{n} H(c_i)}. \tag{18}$$

Note that $I_r = 1$ when the c_i are fully redundant (e.g., identical) and $I_r = 0$ when they are independent.

Figure 10 shows the relative mutual information between pairs of coefficients in the three considered domains as a function of the distance between them. As expected, global linear ICA achieves a large reduction in relative mutual information with regard to the values in the spatial

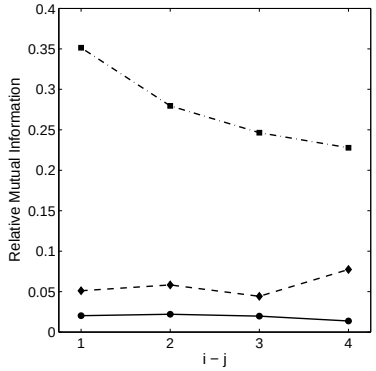


Figure 10: Relative mutual information of pairs of coefficients, (i, j), in the three considered domains as a function of the distance between coefficients, $i - j$. Spatial domain ($I_r(x_i, x_j)$, dash-dot), global ICA ($I_r(c_i, c_j)$, dashed), the proposed local linear ICA ($I_r(r_i, r_j)$, solid).

domain. However the proposed representation achieves an even bigger reduction revealing a weaker statistical dependence between the local-to-global non-linear ICA coefficients.

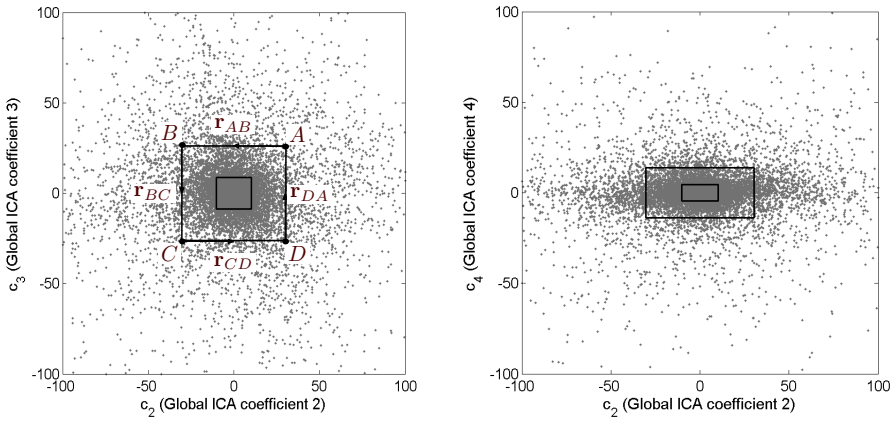
These two facts (removing bow-tie dependencies and mutual information reduction) show that the proposed method increases the independence of the coefficients obtained by a standard global linear ICA technique.

B Independence of the integration path

In this appendix we show that the integral in eq. 7 is independent of the integration path. The proof is based in the standard Stokes' theorem used in classical electrodynamics (Jackson, 1998).

The proposed local linear ICA associates to each point, $\mathbf{x_0}$, a set of vectors, $\mathbf{W}(\mathbf{x_0})$. Thus eq. 7 can be seen as the integration of a vector field. The Stokes' theorem states that the integration of a vector field is independent of the integration path if the integral in closed paths vanishes (Jackson, 1998).

The numerical results below show that in our case, integration in several closed paths in different planes is almost zero. Figure 11 shows two examples of the selected planes defined in the global ICA representation. For each plane, the behavior of the non-linear response in two closed integration paths has been explored. The difference between these two integration paths is that the first one (closer to the origin) goes across low contrast images in highly populated regions while in the second one the contrast is bigger and we have less image samples per unit of volume. The



a) b)

Figure 11: Examples of integration paths in different planes of the image space.

response across these rectangular integration paths can be computed by adding the responses in each segment of the path, $r_{ABCDA} = r_{AB} + r_{BC} + r_{CD} + r_{DA}$, as illustrated in figure 11.a.

The vectors obtained for each segment in the particular path highlighted in figure 11.a, defined in the plane (c_2, c_3), are,

$$r_{AB} = \begin{bmatrix} -0.43 \\ -2.20 \\ -1.10 \\ 0.01 \\ -0.18 \\ -0.39 \end{bmatrix} \quad r_{CD} = \begin{bmatrix} 0.59 \\ 2.00 \\ 1.10 \\ -0.05 \\ 0.19 \\ 0.33 \end{bmatrix} \quad r_{BC} = \begin{bmatrix} -0.52 \\ 1.40 \\ -1.50 \\ 0.33 \\ 0.25 \\ -0.69 \end{bmatrix} \quad r_{DA} = \begin{bmatrix} 0.47 \\ -1.40 \\ 1.60 \\ -0.07 \\ -0.45 \\ 0.56 \end{bmatrix} \tag{19}$$

Here we can see a number of interesting features of the non-linear response: (1) when moving along the AB direction (reduction of the c_2 component) the biggest activity is obtained in the second sensor r_{AB_2}, (2) however, this is not the only active sensor, revealing cross coefficient interactions in the non-linear responses, (3) when moving in the opposite direction, CD, the response vector approximately reverses, $r_{CD} \approx -r_{AB}$. These trends also hold for the other two segments in such a way that the final sum approximately cancels.

In order to show that these responses are close to zero we measured the norm of the response

across the closed path relative to the sum of the norms along the segments:

$$n = \frac{|r_{ABCDA}|}{|r_{AB}| + |r_{BC}| + |r_{CD}| + |r_{DA}|} \tag{20}$$

Table 1 shows the results of this relative norm for the different considered planes and integration paths.

	(c_2, c_4)	(c_2, c_3)	(c_3, c_4)	(c_3, c_5)	(c_4, c_6)
Low Contr.	0.04	0.05	0.04	0.03	0.10
High Contr.	0.12	0.10	0.09	0.05	0.09

Table 1: Relative norms of the responses in the closed paths

These results show that the norm of the response in closed paths is small (about 4%-10% of the average response norm). This suggests that the result of the integral 7 is fairly independent of the integration path, so there is no need to specify the integration path.

REFERENCES

Atick, J. (1992). Could Information Theory provide an ecological theory of sensory processing? *Network: Comp. Neur. Syst.*, 3(2):213–251.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W., editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA.

Barlow, H. B. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241–253.

Cover, T. and Tomas, J. (1991). *Elements of Information Theory*. John Wiley & Sons, New York.

Daugman, J. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394.

Foley, J. (1994). Human luminance pattern mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, 11(6):1710–1719.

Gersho, A. and Gray, R. (1992). *Vector Quantization and Signal Compression*. Kluwer Academic Press, Boston.

Harvey, L. and Doan, V. (1990). Visual masking at different polar angles in the two-dimensional fourier plane. *Journal of the Optical Society of America A*, 7(1):116–127.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–198.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neur. Nets.*, 10(3):626–634.

Hyvärinen, A., Hurri, J., and Väyrynen, J. (2003). Bubbles: a unifying framework for low level statistical properties of natural images sequences. *JOSA A*, 20(7):1237–1252.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons, New York.

Jackson, J. D. (1998). *Classical Electrodynamics*. John Wiley and Sons, NY.

Jutten, C. and Karhunen, J. (2003). Advances in nonlinear blind source separation. *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 245–256.

Kambhatla, N. and Leen, T. (1997). Dimension reduction by local PCA. *Neural Computation*, (9):1493–1500.

Karhunen, J., Malaroiu, S., and Ilmoniemi, M. (2000). Local linear independent component analysis based on clustering. *Intl. J. Neur. Syst.*, 10(6).

Legge, G. (1981). A power law for contrast discrimination. *Vision Research*, 18:68–91.

Legge, G. and Foley, J. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70:1458–1471.

Lin, J. (1999). Factorizing probability density functions: Generalizing ICA. *Proc. of the First Intl. Workshop on ICA and Signal Separation*, pages 313–318.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):127–135.

Malo, J., Epifanio, I., Navarro, R., and Simoncelli, E. (2005). Non-linear image representation for efficient perceptual coding. *To appear in: IEEE Transactions on Image Processing*.

Malo, J., Ferri, F., Albert, J., J.Soret, and Artigas, J. (2000). The role of perceptual contrast non-linearities in image transform coding. *Image & Vision Computing*, 18(3):233–246.

Malo, J., Gutiérrez, J., and Rovira, J. (2004). Perturbation analysis of the changes in V1 receptive fields due to context. *GRC Sensory Coding and the Natural Environment*.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Schwartz, O. and Simoncelli, E. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825.

Simoncelli, E. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, 13:144–149.

Simoncelli, E. and Olshausen, B. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, 24:1193–1216.

van Hateren, J. and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc.R.Soc.Lond. B*, 265:359–366.

Watson, A. (1983). Detection and recognition of simple spatial forms. In Braddick, O. and Sleigh, A., editors, *Physical and Biological Processing of Images*, volume 11 of *Springer Series on Information Sciences*, pages 100–114, Berlin. Springer Verlag.

Watson, A. (1987). Efficiency of a model human image code. *Journal of Optical Society of America A*, 4(12):2401–2417.

Watson, A. and Solomon, J. (1997). A model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14:2379–2391.

B. TRABAJO 2

Importance of quantizer design compared to optimal multigrid motion estimation in video coding.

Jesús Malo, Francesc J. Ferri, **Juan Gutiérrez** e Irene Epifanio.

IEE Electronics Letters, Vol. 36, 9, pp. 807-809, 2000

ELECTRONICS LETTERS

AN INTERNATIONAL PUBLICATION

Inside this issue

27th April 2000 Vol. 36 No. 9

(full contents list inside)

pages 775–856

THE INSTITUTION OF ELECTRICAL ENGINEERS

ELECTRONICS LETTERS

AN INTERNATIONAL PUBLICATION

27 APRIL 2000
ELLEAK 36 (9)

VOLUME 36
775 – 856

NUMBER 9
ISSN 0013-5194

ELECTRONICS LETTERS (ISSN 0013-5194) is published every other week except for one issue in December (total of 25 issues). 2000 annual subscription price £665.00. Single copy £27.00. Air mail service direct to subscribers an additional £62.00.

All subscription inquiries and orders should be sent to IEE Publication Sales, PO Box 96, Stevenage SG1 2SD, United Kingdom.
All other correspondence, including advertisements, should be sent to the IEE Publishing Department.
The advantages of rapid publication given by *Electronics Letters* are nullified if subscribers are subject to excessive delays in the receipt of the journal. The Executive Editor would be grateful to learn of any difficulties subscribers may experience.

Electronics Letters Online

Electronics Letters is available online through the IEE Online Journals service on the IEE's World-Wide Web server. Articles are stored in PDF format and may be viewed and printed using the Adobe® Acrobat® Reader. Features include: browsable table of contents pages for each journal issue, cross-journal searching of bibliographic records (inc. authors, titles, abstracts), and the ability to print articles locally. The issues are added in advance of print publication. For further information, please look at http://ioj.iee.org.uk/ or contact the Marketing Department at Stevenage.

Scope and Guide to Authors

Electronics Letters embraces the entire field of modern electronics, including electronic science and engineering, telecommunications, optoelectronics, and optical communication. *Electronics Letters* specialises in the rapid publication of short international research papers at the cutting edge of electronics technology.
Contributions should be in accordance with the Guide to Authors, which is to be found on the inside back cover or the last page of each issue, and should be addressed as indicated in the Guide.

The Institution of Electrical Engineers
Savoy Place
London WC2R 0BL
United Kingdom

Telephone	+44 (0)171 240 1871
Facsimile	+44 (0)171 240 7735
WWW	http://www.iee.org.uk

Publishing Department
IEE
Michael Faraday House
Six Hills Way
Stevenage SG1 2AY
United Kingdom

Telephone	+44 (0)1438 313311
Facsimile:	
Sales	+44 (0)1438 742792
Editorial	+44 (0)1438 742849
Advertising	+44 (0)1438 318361
Marketing	+44 (0)1438 742840
E-mail:	
Editorial	eletters@iee.org.uk
Marketing	inspec@iee.org.uk

Typeset in England by
Pindar plc
Ryedale Building
First Floor
60 Piccadilly
York YO1 1WU

Printed in UK by
A.McLay & Co. Limited
Longwood Drive
Forest Farm
Cardiff CF4 7ZB

Contents

pages 775–856

27th April 2000 Vol. 36 No. 9

(*continued overleaf*)

This added accuracy in estimating the minimum distance and the weight distribution of the code leads to the computation of a tighter bound on the BER performance. Fig. 1 shows a comparison between the bounds using the averaging technique and that using the modified technique for the (1,5/7,5/7) turbo code. Two puncturing patterns of the systematic sequence are considered. In the first case, one systematic bit is punctured in a period of two ($M = 1$, $p = 2$), while in the second case one systematic bit is punctured in a period of three ($M = 1$, $p = 3$).

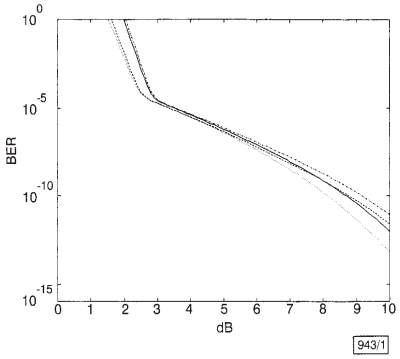


Fig. 1 *Comparison of upper bound on BER using averaging and modified techniques*

— · — averaging [3], $M = 1$, $p = 2$
– – – – averaging [3], $M = 1$, $p = 3$
····· modified [1 1 0]
——— modified [1 0]

The improved tightness of the bound at high values of E_b/N_0 due to the modification is quite evident. The reason behind this can be explained as follows. Over this range of E_b/N_0, the code performance is dominated by the minimum distance of the code. The averaging method estimates the minimum distance of the code to be 4, while the actual minimum distance of the code, as learned from the modified technique, cannot be less than 6. The averaging method assumes the presence of all possible combinations of the two constituent code outputs with different probabilities. This includes associating low weight codewords from both constituent codes even though they might not exist.

There is another significant advantage of the modified technique: it permits the investigation of different puncturing patterns when systematic bits are involved. The averaging technique does not have this freedom, and as long as the puncturing patterns have the same p and M, then the same weight distribution will be obtained. In addition, the modified technique helps in investigating the effect of puncturing systematic symbols compared to puncturing parity symbols, which is also not possible using the averaging technique.

Conclusion: A modified technique for evaluating the weight distribution for punctured systematic bits has been shown to result in a better estimation of the minimum distance, a more accurate evaluation of the weight distribution and a tighter BER performance curve of the code compared to the averaging technique. Also, the modified technique has an added analytical value as it enables the investigation of different patterns of puncturing systematic bits.

Acknowledgment: The authors acknowledge KFUPM for supporting this research.

26 January 2000
Electronics Letters Online No: 20000478
DOI: 10.1049/el:20000478

A.H. Mugaibel and M.A. Kousa (*EE Department, King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, PO Box 1721, Saudi Arabia*)

E-mail: mugaibel@kfupm.edu.sa

References

1 BERROU, C., GLAVIEUX, A., and THITIMAJHIMA, P.: 'Near Shannon limit error correcting coding and decoding : turbo-codes'. Proc. ICC '93, Geneva, May 1993, pp. 1064–1070
2 BERROU, C., and GLAVIEUX, A.: 'Near optimum error correcting coding and decoding: Turbo-codes', *IEEE Trans.*, 1996, **COM-44**, pp. 1261–1271
3 BENEDETTO, S., and MONTORSI, G.: 'Unveiling turbo-codes: some results on parallel concatenated coding schemes', *IEEE Trans.*, 1996, **IT-42**, (2), pp. 409–428

Importance of quantiser design compared to optimal multigrid motion estimation in video coding

J. Malo, F.J. Ferri, J. Gutierrez and I. Epifanio

Adaptive flow computation and DCT quantisation play complementary roles in motion compensated video coding schemes. Since the introduction of the intuitive entropy-constrained motion estimation of Dufaux *et al.*, several optimal variable-size block matching algorithms have been proposed. Many of these approaches put forward their intrinsic optimality, but the corresponding visual effect has not been explored. The relative importance of optimal multigrid motion estimation with regard to quantisation is addressed in the context of MPEG-like coding. It is shown that while simpler (suboptimal) motion estimates give subjective results as good as the optimal motion estimates, small enhancements in the quantiser have significant visual effects. This suggests that more attention should be paid to the quantiser design.

Introduction: In H.263 and MPEG video coders, the original sequence is split into two lower complexity signals that carry information about motion and prediction errors. These are referred to as the displacement vector field (DVF) and the displaced frame difference (DFD). A more complex and accurate DVF gives rise to a better prediction (lower complexity DFD). However, the relative effectiveness of such a detailed description relies heavily on the DCT quantiser used. In other words, the better the quantiser, the less significant the improvement due to motion estimation.

The so-called entropy-constrained multigrid approach [1] explicitly makes use of this trade-off between motion estimation and DFD complexity. Since this approach was introduced, great effort has been devoted to obtaining analytical [2 – 4] or numerical [5] optimal entropy-constrained quadtree DVF decompositions. These approaches criticise the (faster) entropy measure of the DFD in the spatial domain of Dufaux *et al.* because it does not take into account the effect of the selective DCT quantiser. This necessarily implies a suboptimal bit allocation between DVF and DFD. The literature [2 – 5] reports the quantitative optimality of the proposed methods, but the practical (subjective) effect of this gain on the reconstructed sequence is not analysed. In particular, only (perceptually unweighted) SNR or MSE distortion measures are given and no explicit comparison of the decoded sequences is shown.

In this Letter, the relative importance of an optimal adaptive motion estimation is addressed. Combinations of variable-size block matching algorithms (VSBMAs) [1, 5] and different quantisation schemes [6 – 8] are evaluated. Distortion results with perceptually meaningful measures [9, 10] and explicit comparisons of the reconstructed frames are shown.

Algorithms and distortion measures: Two different approaches can be followed to design VSBMAs. First, the final bit-rate can be optimised taking into account both adaptive motion and error encoding [2 – 4], given a particular coding scheme. Second, a splitting criterion which is related to the entropy of both motion and error signals can be used for adaptive motion estimation [1, 5]. In the latter case, the corresponding VSBMAs locally increase the resolution if the volume of the resulting signal is reduced, i.e. a block is split if $H(DVF_{split}) + H(DFD_{split}) < H(DVF_{nosplit}) +$

$H(DFD_{nosplit})$, where $H(DVF)$ is the entropy of the vector field plus the quadtree, and $H(DFD)$ is a measure of the complexity of the residual error signal. The entropy $H(DFD)$ can be computed either directly from the DFD in the spatial domain (suboptimal) [1] or as the entropy of the quantised transform of the DFD (optimal) [5]. In the optimal case, motion estimation is directly related to coding effectiveness. The difference between this and other optimal methods [2 – 4] is in the way this optimisation is stated and performed. However, the best algorithms of each approach are optimal because the particular features of the quantiser are taken into account.

Any VSBMA scheme can be applied using different quantisers. All the DCT quantisers considered here, [6] as well as [7, 8], are based on a perceptually uniform quantisation of the DCT to restrict the maximum perceptual error for every possible input. It has been shown [7] that, under a simple linear perception model, the general equations that define the perceptually uniform quantiser reduce to a simple JPEG-like quantiser based on the contrast sensitivity function (CSF) [6]. If a more accurate nonlinear perception model is used instead, a different quantiser which gives better subjective results on still images [7, 8] is obtained (hereafter referred to as 'enhanced').

The use of meaningful distortion measures in video coding design is not a simple question [9]. The widely known dependence of human visual sensitivity on spatial frequency makes a fundamental difference between plain Euclidean distortion metrics (e.g. MSE or SNR) and any (even the simplest) perceptually weighted metric. For this Letter, distortion was measured with a nonlinear perceptual metric [10] in order to quantify the subjective results obtained by the approaches considered. Simpler [10] or more complex perceptual metrics [9] would lead to the same qualitative results.

Table 1: Relative volume of the DVF

	TAXI	RUBIK	YOSEMITE	TREES	Average
					%
FSBMA	12.57	21.71	46.19	39.21	30 ± 8
Subopt. VSBMA	1.89	3.29	8.69	4.97	4.7 ± 1.5
Optimal VSBMA	0.94	1.87	4.32	3.24	2.6 ± 0.7

Results and discussion: The four possible combinations using optimal and suboptimal VSBMAs and the two DCT quantisers mentioned above were tried on several standard monochrome QCIF sequences at 200kbit/s. To highlight the relative differences among the approaches considered, only the first frame was intracoded, and only forward prediction was used in the remaining frames. As a consequence, the results may seem abnormally distorted at this rate. Both Euclidean and perceptual MSE distortions were computed at each frame in all cases. Also, resulting volumes for DVF and DFD signals were obtained. Specifically, Table 1 shows the relative volume of the motion flow with regard to the total volume of the encoded signal using a JPEG-like quantiser and different motion estimates including fixed-size BMA (FSBMA) as a reference. These and the corresponding results using an enhanced quantiser show a dramatic decrease in the DVF volume of both VSBMAs with regard to FSBMA but, at the same time, an almost negligible difference between the two VSBMA approaches.

Fig. 1 shows the distortion in the reconstructed sequence (RUBIK) computed at each decoded frame. Fig. 2 shows the seventh reconstructed frame using the four different approaches considered. The results in Table 1 and Fig. 1 do indeed confirm the quantitative advantages of optimal motion estimation: the flow volume and the distortion (no matter the measure) are consistently reduced. However, this diminution is almost negligible with regard to the benefits of a better quantisation. This can be numerically assessed by looking at the differences in the corresponding curves in Fig. 1 and, more importantly, by looking at the corresponding reconstructed sequences.

In this case, the simple increase in the quantiser bandpass that comes from considering the visual nonlinearities [7, 8, 10] may explain the enhancement. Owing to the noisy (high frequency) nature of the error signal, wide-band quantisers may be better than narrower-band (CSF-based) quantisers. In fact, the benefits of this enhanced quantiser are more apparent in video frames (Fig. 2) than in still images [7, 8]. The significant improvement due to a particular enhancement of the quantiser suggests that quantiser design may be more important than optimal motion estimation.

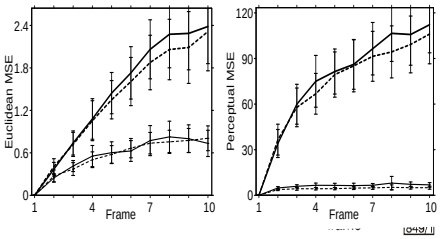


Fig. 1 *MSE measures*

JPEG-like quantiser, suboptimal VSBMA
enhanced quantiser, suboptimal VSBMA
JPEG-like quantiser, optimal VSBMA
enhanced quantiser, optimal VSBMA

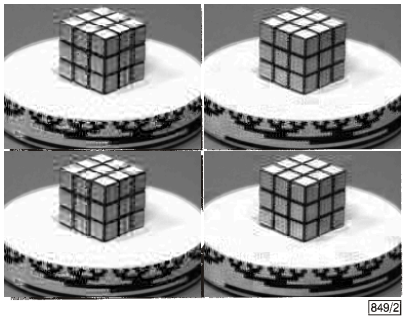


Fig. 2 *Reconstructed frames*

Upper and lower frames correspond to suboptimal and optimal VSBMA, respectively
Left and right frames correspond to JPEG-like and enhanced quantiser, respectively

In contrast, it has been reported [5] that a more robust and meaningful motion estimate is obtained as a side effect of the optimal interaction between the motion estimation and a perceptual quantiser (due to its bandpass nature). As can be seen by comparing the upper and lower frames in Fig. 2, this advantage in motion estimation has negligible visual effects on the reconstructed sequence. Faster suboptimal multigrid methods are good enough from the visual perspective. However, if this increased robustness effect is exhaustively confirmed, it may be of interest to higher-level uses of motion information such as in model-based video coding.

Acknowledgment: This work has been supported by CICYT (TIC) projects 1FD97-0279 and 1FD97-1910

14 March 2000
Electronics Letters Online No: 20000645
DOI: 10.1049/el:20000645

J. Malo (*Dept. d'Òptica, Universitat de València, Dr. Moliner, 50, 46100 Burjassot, València, Spain*)

E-mail: jesus.malo@uv.es

F.J. Ferri, J. Gutierrez and I. Epifanio (*Dept. d'Informàtica, Universitat de València, Dr. Moliner, 50, 46100 Burjassot, València, Spain*)

References

1 DUFAUX, F., and MOSCHENI, F.: 'Motion estimation techniques for digital TV: A review and new contribution', *Proc. IEEE*, 1995, **83**, (6), pp. 858–876

2 SCHUSTER, G.M., and KATSAGGELOS, A.K.: 'Rate-distortion based video compression' (Kluwer Acad. Publ., Boston, 1997)

3 SCHUSTER, G.M., and KATSAGGELOS, A.K.: 'A video compression scheme with optimal bit allocation among segmentation, motion and residual error', *IEEE Trans. Image Process.*, 1997, **6**, (11), pp. 1487–1502
4 LEE, J.: 'Joint optimization of block size and quantization for quadtree-based motion estimation', *IEEE Trans. Image Process.*, 1998, **7**, (6), pp. 909–912
5 MALO, J., FERRI, F., ALBERT, J., and ARTIGAS, J.M.: 'Splitting criterion for hierarchical motion estimation based on perceptual coding', *Electron. Lett.*, 1998, **34**, (6), pp. 541–543
6 WALLACE, G.K.: 'The JPEG still picture compression standard', *Comm. ACM*, 1991, **34**, (4) pp. 30–44
7 MALO, J., FERRI, F., ALBERT, J., and SORET, J.: 'Comparison of perceptually uniform quantisation with average error minimization in image transform coding', *Electron. Lett.*, 1999, **35**, (13), pp. 1067–1068
8 MALO, J., FERRI, F., ALBERT, J., SORET, J., and ARTIGAS, J.: 'The role of perceptual contrast non-linearities in image transform quantization', *Image Vis. Comput.*, 2000, **18**, (3), pp. 233–246
9 WINKLER, S.: 'Issues in vision modeling for perceptual video quality assessment', *Signal Process.*, 1999, **78**, pp. 231–252
10 MALO, J., PONS, A.M., and ARTIGAS, J.M.: 'Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain', *Image Vis. Comput.*, 1997, **15**, pp. 535–548

Investigation into decoding algorithm for turbo codes

D. Raphaeli and A. Gurevitz

An investigation into the turbo decoder is presented. It was carried out to determine whether the decoded bit sequence constitutes a (at least) local maximum in the likelihood between possible codewords in the composite code trellis. The answer was obtained experimentally using a modified iterative turbo decoder. Results show that the turbo decoder does not necessarily lead to a maximum likelihood sequence estimator being realised for the composite code. Moreover, finding a closer code in a Euclidean distance sense only degrades the bit error rate of the system.

Introduction: Turbo codes were first introduced in [1] and since then have been subject to a large amount of interest, and consequently have been widely investigated by the coding community. In this Letter we present an investigation into the nature of detected codewords derived by a turbo decoder in relation to the (maximum or minimum) Euclidean distance between the received and detected sequences.

The usual method of decoding convolutional codes involves using a maximum likelihood sequence estimator (MLSE), which finds the sequence in the trellis that is nearest in Euclidean distance to the transmitted sequence. In other words, it finds the most probable sequence and, consequently, produces the lowest probability of block errors. This does not mean however that the bit error probability (BER) is minimised.

Although the exact nature of the iterative decoding process is not known, it is natural to believe that iterative decoding is a process which seeks to obtain maximum likelihood in an iterative way, just like the EM algorithm [2], for example. In the following we will show that this is not true, since the point of convergence of the decoder is not ML, even in a local sense.

At the end of the iterative process, after the decoder has reached a final sequence of detected bits, we attempt to improve the performance of the decoder by inverting some bits that decrease the Euclidean distance, and thus hopefully decrease the block error rate or bit error rate of the receiver. We continue inverting the bits until the minimum, at least local, of the Euclidean distance is reached. The results lead to two conclusions. The first is that the sequence at the end of the iterative process is not a local minimum of the Euclidean distance, and hence is not a maximum likelihood process. The second is that obtaining a higher likelihood for the sequence degrades the BER instead of improving it. Erom this we can postulate that the iterative decoder in fact approximates the minimum BER decoder, or a maximum *a posteriori* (MAP) decoder, which has maximum likelihood per bit instead of per sequence.

The system investigated in this Letter comprises a turbo code that is formed by the concatenation of two recursive systematic codes, linked together by non-uniform interleaving [1]. A turbo decoder uses the MAP algorithm for each sub-code [3] to approximate the *a posteriori* probabilities of the different transmitted bits.

(i) *MLSE detector*: It has been shown in [4] that the MLSE for convolutional codes minimises the probability of block errors for the sequence of information bits. The MLSE detector searches for the minimum Euclidean distance path through the trellis that characterises the memory of the transmitted signal. Consider for example a coded sequence that is transmitted by binary coherent PSK symbols:

$$\mathbf{s}^{(m)} = \left\{ s_1^{(m)}, s_2^{(m)}, s_3^{(m)}, ..., s_k^{(m)} \right\} \tag{1}$$

in which each symbol $s_k^{(m)}$ can represent either two points in the constellation $\pm\sqrt{\varepsilon_c}$ where ε_c is the transmitted signal energy for each coded bit. Assuming a memoryless additive white Gaussian noise (AWGN) channel and soft decision decoding, the output of the matched filter demodulator received sequence would be r_1, r_2, ..., r_k where

$$r_k = \pm\sqrt{\varepsilon_c} + n_k \tag{2}$$

and n_k is a zero mean white Gaussian random variable with variance $\sigma_n^2 = N_0/2$. In this case the MLSE detector would determine $\mathbf{s}^{(m)}$ that maximises the conditional probability density function (PDF):

$$p(r_1, r_2, ..., r_k|\mathbf{s}^{(m)}) = \prod_{k=1}^{K} p\left(r_k|s_k^{(m)}\right)$$
$$= \left(\frac{1}{\sqrt{2\pi}\sigma_n}\right)^K \exp\left[-\sum_{k=1}^{K}\frac{(r_k - s_k^{(m)})^2}{2\sigma_n^2}\right] \tag{3}$$

By taking the logarithm of eqn. 3 we find that an equivalent ML sequence detector selects the sequence $\mathbf{s}^{(m)}$ that minimises the Euclidean distance metric:

$$\sum_{k=1}^{K}\left(r_k - s_k^{(m)}\right)^2 \tag{4}$$

(ii) *MAP detector*: In contrast to the maximum-likelihood sequence detector, the MAP detector makes symbol-by-symbol decisions based on the computation of the maximum *a posteriori* probability for each detected symbol. Therefore this detector is optimum in the sense that it minimises the probability of a symbol error. Consider a coded symbol sequence starting at time 1 and ending at time T:

$$\mathbf{s}^{(m)} = \left\{ s_1^{(m)}, s_2^{(m)}, s_3^{(m)}, ..., s_T^{(m)} \right\} \tag{5}$$

The MAP algorithm would determine the symbol $s_t^{(m)}$ that maximises the conditional PDF:

$$p\left(s_t^{(m)}|r_1, r_2, ..., r_T\right) \tag{6}$$

for each symbol in the sequence. Hence it will achieve the optimum symbol error rate. Turbo codes can be described as composite convolutional codes. Consequently, the detected symbols constitute a path in the composite trellis. Our objective was to find whether this path is a minimum Euclidean distance path through the trellis and whether the sequence or block error rate would improve by finding a nearer path.

System model: Consider that binary source bits $\{d_i\}$ are encoded by a turbo encoder, similar to the encoder used in [1]. The turbo encoder is formed from two identical recursive encoders, separated by a random interleaver (π). The coded bits are BPSK modulated and transmitted over an AWGN channel. The BPSK output symbols will be referred to as $\{c_k\}$.

The general decoder structure is presented in Fig. 1. The turbo decoder shown is based on two MAP blocks [5]. Each of them produces an LLR of the source data and code bits. At the end of the turbo decoder iterations, a hard decision on the output bit LLR will provide an estimation of the transmitted bit sequence $\hat{d}_i$.

C. TRABAJO 3

Perceptually weighted optical flow for motion-based segmentation in MPEG-4.

Jesus Malo, **Juan Gutiérrez**, Irene Epifanio y Francesc J. Ferri.

IEE Electronics Letters, Vol 36, No. 20, pp. 1693-1694, 2000

ELECTRONICS LETTERS

AN INTERNATIONAL PUBLICATION

Inside this issue

(full contents list inside)

28th September 2000 Vol. 36 No. 20

pages 1675–1752

IEE

THE INSTITUTION OF ELECTRICAL ENGINEERS

ELECTRONICS LETTERS

AN INTERNATIONAL PUBLICATION

28 SEPTEMBER 2000
ELLEAK 36 (20)

VOLUME 36
1675 – 1752

NUMBER 20
ISSN 0013-5194

ELECTRONICS LETTERS (ISSN 0013-5194) is published every other week except for one issue in December (total of 25 issues). 2000 annual subscription price £665.00. Single copy £27.00. Air mail service direct to subscribers an additional £62.00.

All subscription inquiries and orders should be sent to IEE Publication Sales, PO Box 96, Stevenage SG1 2SD, United Kingdom.
All other correspondence, including advertisements, should be sent to the IEE Publishing Department.
The advantages of rapid publication given by *Electronics Letters* are nullified if subscribers are subject to excessive delays in the receipt of the journal. The Executive Editor would be grateful to learn of any difficulties subscribers may experience.

Electronics Letters Online

Electronics Letters is available online through the IEE Online Journals service on the IEE's World-Wide Web server. Articles are stored in PDF format and may be viewed and printed using the Adobe® Acrobat® Reader. Features include: browsable table of contents pages for each journal issue, cross-journal searching of bibliographic records (inc. authors, titles, abstracts), and the ability to print articles locally. The issues are added in advance of print publication. For further information, please look at http://ioj.iee.org.uk/ or contact the Marketing Department at Stevenage.

Scope and Guide to Authors

Electronics Letters embraces the entire field of modern electronics, including electronic science and engineering, telecommunications, optoelectronics, and optical communication. *Electronics Letters* specialises in the rapid publication of short international research papers at the cutting edge of electronics technology.
Contributions should be in accordance with the Guide to Authors, which is to be found on the inside back cover or the last page of each issue, and should be addressed as indicated in the Guide.

The Institution of Electrical Engineers
Savoy Place
London WC2R 0BL
United Kingdom

Telephone	+44 (0)171 240 1871
Facsimile	+44 (0)171 240 7735
WWW	http://www.iee.org.uk

Publishing Department
IEE
Michael Faraday House
Six Hills Way
Stevenage SG1 2AY
United Kingdom

Telephone	+44 (0)1438 313311
Facsimile:	
Sales	+44 (0)1438 742792
Editorial	+44 (0)1438 742849
Advertising	+44 (0)1438 318361
Marketing	+44 (0)1438 742840
E-mail:	
Editorial	eletters@iee.org.uk
Marketing	inspec@iee.org.uk

Chief Executive of the IEE
Alf Roberts, PhD, CEng, FIEE, FInsP

Editorial Staff
Gill Wheeler (Managing Editor)
Anoushka Kulikowski
Stuart Govan
Harry Colson

HONORARY EDITORS
Sir Eric Ash, CBE, FREng, FRS
Prof. Peter Clarricoats, CBE, FREng, FRS
Prof. Chris Toumazou

Typeset in England by
Pindar plc
Ryedale Building
First Floor
60 Piccadilly
York YO1 1WU

Printed in UK by
A.McLay & Co. Limited
Longwood Drive
Forest Farm
Cardiff CF4 7ZB

Contents

pages 1675–1752 28th September 2000 Vol. 36 No. 20

(*continued overleaf*)

2 IRETON, M.A., and XYDEAS, C.S.: 'Classification of shape for content retrieval of images in a multimedia database'. Proc. Sixth Int. Conf. on Digital Processing of Signals in Communications, 1991, pp. 111–116
3 SERRA, J.: 'Image analysis and mathematical morphology' (Academic Press, 1992), vol. 2
4 GOUTSIAS, J.: 'Morphological representation of discrete and binary images', *IEEE Trans. Signal Process.*, 1991, **39**, (6), pp. 1369–1379
5 SOILLE, P.: 'Morphological image analysis. Principles and applications' (Springer, 1999)

Perceptually weighted optical flow for motion-based segmentation in MPEG-4 paradigm

J. Malo, J. Gutierrez, I. Epifanio and F.J. Ferri

In the MPEG-4 paradigm, the sequence must be described in terms of meaningful objects. This meaningful, high-level, representation should emerge from low-level primitives such as optical flow and prediction error which are the basic elements of previous-generation video coders. The accuracy of the high-level models strongly depends on the robustness of the primitives used. It is shown how perceptual weighting in optical flow computation gives rise to better motion estimates which consistently improve motion-based segmentation compared to equivalent unweighted motion estimates.

Introduction: MPEG-4 recommendations stress the relevance of high-level sequence representation [1]. Motion-based segmentation algorithms are thus crucial for identifying meaningful regions and for obtaining models to predict their evolution. In this context, it is interesting to analyse whether the low-level motion information used in previous-generation video coders may be useful as a starting point for the segmentation. This is not a trivial question because the block matching algorithms (BMAs) which are used in MPEG-1,2 or H.263 for motion estimation were not intended to give a qualitatively meaningful flow, but a good prediction tool in a rate-distortion sense [2].

A great effort has been devoted to optimising multigrid locally adaptive BMA for its use in H.263-like (low-level) coders [2 – 5]. The entropy constraint for multigrid motion estimation was originally introduced in [2] to obtain a balanced bit allocation between motion information and prediction errors. Most of the posterior works in adaptive BMA including the perceptual DCT quantisation have been motivated by purely quantitative rate-distortion reasoning [3, 4]. Beyond the optimal rate-distortion result, which has been shown to be visually negligible [6], the bandpass nature of the perceptual quantiser introduces desirable properties on the flow [5]. The coherence and robustness of this flow (which could make it suitable for higher level purposes) appears to be of capital importance compared to its optimality.

The aim of this Letter is to point out that perceptual weighting in the flow computation makes the (low-level) sparse H.263 motion information robust enough to be used for MPEG4-like coding. To test this hypothesis, a well-known motion-based segmentation algorithm for high-level video coding [7, 8] was used with perceptually weighted and perceptually unweighted BMAS used in low-level H.263-like coders.

Perceptually weighted multigrid BMA: The general idea behind BMAs is the search of the block displacement which minimises the prediction error. Multigrid BMA is a coarse-to-fine procedure in which the resolution is locally increased by splitting the blocks used in the displacement search. The splitting criterion determines where to focus the motion estimation effort. According to entropy-constrained motion estimation, the resolution of the optical flow should be increased in a particular position if the volume of the final signal is reduced [2, 5]. In particular, a block should be split if the sum of the entropies of the motion flow and the prediction error or displacement frame difference, $H(\vec{D}) + H(DFD)$, is reduced. The entropy $H(DFD)$ can be computed either in the spatial domain [2] or in the frequency domain after the application of the perceptual bandpass DCT quantiser [5]. The latter option is optimal because it takes into account the actual volume of the encoded signal. Beyond this quantitative difference, while the first unweighted criterion operates in the same way for every resolution, the second perceptually weighted criterion behaves in a scale-dependent way. This scale-dependent behaviour can be explained in terms of the uncertainty relation: as the spatial support of a signal, Δx, and its bandwidth, Δf, are related by $\Delta x \cdot \Delta f = k$, the spectrum of the prediction errors depends on the resolution. Large spatial support in low resolution implies low frequency content errors. Small spatial support in high resolution spreads the error spectrum. In this context, the low bandpass nature of the perceptual DCT quantiser gives high entropy values for low resolution DFDs and low entropy values for high resolution DFDs. As a result, while a permissive splitting criterion is obtained at low resolution, a restrictive criterion is obtained at high resolution. This scale-dependent strategy gives rise to a more robust flow which is focused on perceptually relevant motions [5]. This qualitative difference of the perceptually weighted flow should improve the results of the motion-based segmentation.

Motion-based segmentation and its evaluation: The algorithms to obtain a layered representation of a video sequence [7, 8] use a displacement field as input low-level motion information. From this flow, an arbitrary initial segmentation is iteratively refined estimating an affine motion model in each region and merging the regions with similar affine parameters. Given a pair of frames from a sequence with n moving objects, the segmentation algorithm gives m masks, representing the identified objects and the corresponding affine models.

To assess the relative performance of the flows under consideration, the following dissimilarity measure between the (manually segmented) object O_i and the mask M (which maximises $M_j \cap O_i$ for all the obtained masks M_j) was used:

$$\varepsilon_i = \int_{M \cap O_i^C} D_i(x)dx + \int_{O_i \cap M^C} D_i(x)dx \qquad (1)$$

where C stands for complementary, and the chamfer distance, $D_i(x)$, to the object O_i [9] progressively penalises unmatched pixels which are far from the object boundary. The segmentation error for a frame, ε, is the sum of all ε_i, It has been empirically found that this error measure adequately describes the quality of the segmentation.

Results and discussion: Several experiments were carried out using the referred layer identification algorithm [7, 8] which was initialised with three different sparse flows (one motion vector per block): fixed-size BMA, unweighted variable-size BMA [2], and perceptual variable-size BMA [5]. Results regarding two standard sequences, TAXI and RUBIK, are considered here because they represent extreme cases for the segmentation algorithm. Small moving objects with simple translations, and one large object rotating around an axis, respectively. While in the first sequence even simple velocity clustering could solve the problem, in the second one, regions with very different velocities have to be merged in a single complex-motion object.

Table 1: Segmentation errors, number of iterations and identified regions

	RUBIK (2 objects)			TAXI (4 objects)		
	ε	Iteratations	Regions	ε	Iteratations	Regions
FSBMA	1.00±0.05	25	21.00±4.8	1.00±0.14	25	15.8±3.8
Unweighted VSBMA	0.46±0.06	21±5	4.4±2.2	0.52±0.15	11±9	5.8±2.0
Perceptual	0.27±0.03	7±6	2.1±0.3	0.49±0.14	4±1	4.3±0.5

Table 1 shows the average segmentation error (normalised by the maximum in each sequence), the number of iterations and the number of identified regions in each case. Fig. 1 shows an example of the segmentations achieved. The block appearance of the layers is obviously due to the sparseness of the input flows (the segmentation algorithm is usually initialised with one motion vector per pixel [7, 8]).

The following trends can be identified from the obtained results: a) the segmentation is very sensitive to the robustness and coherence of the sparse flow, which implies that the block size cannot

be arbitrarily reduced to improve the resolution, and b) the segmentation is better when the blocks are fairly small compared to the size of the moving regions (see the variable-size BMA error results for RUBIK -large object- and TAXI -small objects-).

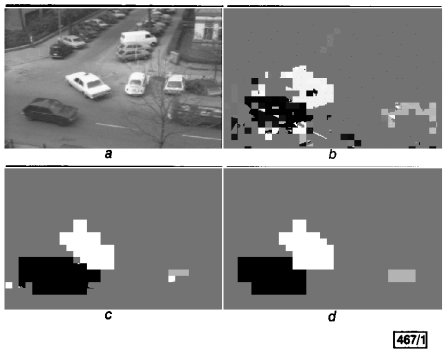


Fig. 1 *Segmentation results*

a TAXI, frame 9
b FSBMA, ε =1.02
c Unweighted VSBMA, ε = 0.49
d Perceptual VSBMA, ε = 0.35

This general behaviour may be explained in terms of the data needed to estimate the six parameters of the affine models. Ideally, a minimum of three independent vectors are needed to segment a region. In real situations, more vectors per region will be needed. Consequently, the segmentations of large objects are comparatively better. However, as the number of motion measurements is reduced when using sparse flows (with regard to dense flows), their robustness becomes critical. This is why the results based on sparse fixed-size BMA are extremely poor. However, if the variable-size BMA used is robust enough, reliable rough segmentations are still possible. In particular, the increased robustness of the perceptually weighted flow speeds up the convergence of the segmentation algorithm and minimises the segmentation error (Table 1). Moreover, a closer estimation of the actual number of moving objects is obtained due to the smoothness of the flow within moving regions, which prevents false splittings.

In conclusion, despite the fact that optimal (perceptual) variable-size BMA [3 – 5] does not improve the quality of the reconstructed sequence in the context of H.263 [6], the results presented here show that perceptual weighting makes this flow suitable for obtaining robust rough segmentations from sparse motion information.

Acknowledgment: The authors wish to thank E.H. Adelson for his helpful comments. This work has been supported by CICYT (TIC) projects 1FD97-0279 and 1FD97-1910.

28 July 2000
Electronics Letters Online No: 20001222
DOI: 10.1049/el:20001222

J. Malo (*Dept. Òptica, Universitat de València Dr. Moliner, 50. 46100 Burjassot, València. Spain*)

E-mail: jesus.malo@uv.es

J. Gutierrez, I. Epifanio and F.J. Ferri (*Dept. d'Informàtica, Universitat de València Dr. Moliner, 50. 46100 Burjassot, València. Spain*)

References

1 EBRAHIMI, T., and HORNE, C.: 'MPEG-4 natural video coding: an overview', *Signal Process. Image Commun.*, 2000, **15**, (4), pp. 365–385
2 DUFAUX, F., and MOSCHENI, F.: 'Motion estimation techniques for digital TV: a review and new contribution', *Proc. IEEE*, 1995, **83**, (6), pp. 858–876
3 SCHUSTER, G.M., and KATSAGGELOS, A.K.: 'A video compression scheme with optimal bit allocation among segmentation, motion and residual error', *IEEE Trans. Image Process.*, 1997, **6**, (11), pp. 1487–1502
4 LEE, J.: 'Joint optimization of block size and quantization for quadtree-based motion estimation', *IEEE Trans. Image Process.*, 1998, **7**, (6), pp. 909–912
5 MALO, J., FERRI, F., ALBERT, J., and ARTIGAS, J.M.: 'Splitting criterion for hierarchical motion estimation based on perceptual coding', *Electron. Lett.*, 1998, **34**, (6), pp. 541–543
6 MALO, J., FERRI, F., GUTIERREZ, J., and EPIFANIO, I.: 'Importance of quantiser design compared to optimal multigrid motion estimation in video coding', *Electron. Lett.*, 2000, **36**, (9), pp. 807–809
7 WANG, J.Y.A., and ADELSON, E.H.: 'Representing moving images with layers', *IEEE Trans. Image Process.*, 1994, **3**, (5), pp. 625–638
8 TORRES, L., GARCIA, D., and MATES, A.: 'A robust motion estimation and segmentation approach to represent moving images with layers'. IEEE Int. Conf. Acoust. Speech Sig. Proc., Munich, Germany, 1997, pp. 2981–2984
9 BORGEFORS, G.: 'Distance transformations in digital images'. CVGIP, 1986, Vol. 34, pp. 344–371

Resolution of problem with use of closing for texture segmentation

E.R. Davies

The morphological closing operation is widely known to be useful for separating regions containing high and low densities of particles. It is shown that the area of the region of high density becomes enlarged as the density of particles in the low density region increases - an effect that also occurs if impulse noise appears. It is found that the size increase is calculable and can approximately be cancelled by application of a suitable additional erosion operation, thereby improving the accuracy of the final representation.

Introduction: Texture analysis is an important area of machine vision, and is relevant not only for segmenting one region of an image from another (as in many remote sensing applications), but also for characterising regions absolutely - as is necessary when performing surface inspection (for example, when assessing the paint finish on automobiles). Many methods have been employed for texture analysis. These range from the widely used grey-level co-occurrence matrix approach to Law's texture energy approach, and from use of Markov random fields to fractal modelling [1]. One of the least computation intensive is Laws' method, which involves application of a number of convolution filters to extract spots, edges, lines, waves, ripples and other microfeatures, and then combines them using smoothing operations.

In fact, there are approaches that involve even less computation and which are applicable when the textures are particularly simple and the shapes of the basic texture elements are not especially critical. For example, if it is required to locate regions containing small objects, simple morphological operations applied to thresholded versions of the image might be appropriate (Fig. 2) [2, 3]. Such approaches can be used for locating regions containing seeds, grains, nails, sand or other materials, either for assessing the overall quantity or spread or for determining whether there are regions which have not yet been covered. The basic operation to be applied is the dilation operation, which combines the individual particles into fully connected regions. This method is suitable not only for connecting individual particles but also for separating regions containing high and low densities of such particles. It is well known that the expansion characteristic of the dilation operation can be largely cancelled by a subsequent erosion operation, using the same morphological kernel. Indeed, if the particles are always convex and well separated, the erosion should exactly cancel the dilation, though in general the combined operation (known as closing) is not a null operation, and this is relied upon in the above connecting operation.

We have applied closing operations to images of cereal grains containing dark rodent droppings in order to consolidate the droppings (which contain significant speckle - and therefore holes when the images are thresholded) and thus to make them more readily recognisable from their shapes. However, the result has been rather unsatisfactory as dark patches on the grains tend to combine with the dark droppings: this has the effect of distorting the shapes and also makes the objects larger. We have been able

D. TRABAJO 4

Perceptual feed-back in multigrid motion estimation using an improved DCT quantization.

Jesús Malo, **Juan Gutiérrez**, Irene Epifanio, Francesc J. Ferri y José María Artigas

IEEE Transactions on Image Processing Vol.10, 10, pp. 1411-1427, 2001

IEEE TRANSACTIONS ON

IMAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

www.ieee.org/sp/index.html

OCTOBER 2001 VOLUME 10 NUMBER 10 IIPRE4 (ISSN 1057–7149)

PAPERS

(Contents Continued on Back Cover)

IEEE TRANSACTIONS ON IMAGE PROCESSING (ISSN 1057-7149) is published monthly by the Institute of Electrical and Electronics Engineers, Inc. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society/Council, or its members. **IEEE Corporate Office:** 3 Park Avenue, 17th Floor, New York, NY 10016-5997. **IEEE Operations Center:** 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. **NJ Telephone:** +1 732 981 0060. **Price/Publication Information:** Individual copies: IEEE Members $10.00 (first copy only), nonmembers $20.00 per copy. (Note: Add $4.00 postage and handling charge to any order from $1.00 to $50.00, including prepaid orders.) Member and nonmember subscription prices available upon request. Available in microfiche and microfilm. **Copyright and Reprint Permissions:** Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use of patrons, provided the per-copy fee indicated in the code at the bottom of the first page is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For all other copying, reprint, or republication permission, write to Copyrights and Permissions Department, IEEE Publications Administration, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331.

Periodicals Postage Paid at New York, NY and at additional mailing offices. **Postmaster:** Send address changes to IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. GST Registration No. 125634188. Printed in U.S.A.

Perceptual Feedback in Multigrid Motion Estimation Using an Improved DCT Quantization

Jesús Malo, Juan Gutiérrez, I. Epifanio, Francesc J. Ferri, and José M. Artigas

***Abstract*—In this paper, a multigrid motion compensation video coder based on the current human visual system (HVS) contrast discrimination models is proposed. A novel procedure for the encoding of the prediction errors has been used. This procedure restricts the maximum perceptual distortion in each transform coefficient. This subjective redundancy removal procedure includes the amplitude nonlinearities and some temporal features of human perception. A perceptually weighted control of the adaptive motion estimation algorithm has also been derived from this model. Perceptual feedback in motion estimation ensures a perceptual balance between the motion estimation effort and the redundancy removal process. The results show that this feedback induces a scale-dependent refinement strategy that gives rise to more robust and meaningful motion estimation, which may facilitate higher level sequence interpretation. Perceptually meaningful distortion measures and the reconstructed frames show the subjective improvements of the proposed scheme versus an H.263 scheme with unweighted motion estimation and MPEG-like quantization.**

***Index Terms*—Entropy constrained motion estimation, nonlinear human vision model, perceptual quantization, video coding.**

I. Introduction

In natural video sequences to be judged by human observers, two kinds of redundancies can be identified: 1) *objective redundancies*, related to the spatio-temporal correlations among the video samples and 2) *subjective redundancies*, which refer to the data that can be safely discarded without perceptual loss. The aim of any video coding scheme is to remove both kinds of redundancy. To achieve this aim, current video coders are based on motion compensation and two-dimensional (2-D) transform coding of the residual error [1]–[4]. The original video signal is split into motion information and prediction errors. These two lower complexity sub-sources of information are usually referred to as displacement vector field (DVF) and displaced frame difference (DFD), respectively.

In the most recent standards, H.263 and MPEG-4 [4], [5], the fixed-resolution motion estimation algorithm used in H.261 and MPEG-1 has been replaced by an adaptive, variable-size block matching algorithm (BMA) to obtain improved motion estimates [6]. Spatial subjective redundancy is commonly reduced through a perceptually weighted quantization of a transform of the DFD. The bit allocation among the transform coefficients is based on the spatial frequency response of simple (linear and threshold) perception models [1]–[3].

In this context, there is a clear tradeoff between the effort devoted to motion compensation and transform redundancy removal. On the one hand, better motion estimation may lead to better predictions and should alleviate the task of the quantizer. On the other hand, better quantization techniques may be able to remove more redundancy, thereby reducing the predictive power needed in the motion estimate. Most of the recent work on motion estimation for video coding has been focused on the adaptation of the motion estimate to a *given quantizer* to obtain an good balance between these elements. Since the introduction of the intuitive (suboptimal) entropy-constrained motion estimation of Dufaux *et al.* [7], [8] several optimal, variable-size BMAs have been proposed [9]–[12]. These approaches put forward their intrinsic optimality, but the corresponding visual effect and the relative importance of the motion improvements versus the quantizer improvements have not been deeply explored, mainly because of their subjective nature.

This paper adresses the problem of the tradeoff between multigrid motion estimation and error quantization in a different way. An improved (nonlinear) perception model inspires the whole design to obtain a coder that preserves no more than the subjectively significant information. The role of the perceptual model in the proposed video coder scheme is twofold. First, it is used to simulate the redundancy removal in the human visual system (HVS) through an appropriate perceptually matched quantizer. Second, this perceptual quantizer is used to control the adaptive motion estimation. This control introduces a perceptual feedback in the motion estimation stage. This perceptual feedback limits the motion estimation effort, avoiding superfluous prediction of details that are perceptually negligible and will be discarded by the quantizer. The bandpass shape of the perceptual constraint to the motion estimation gives a scale-dependent control criterion that may be useful for discriminating between significant and noisy motions. Therefore, the benefits of including the properties of the biological filters in the design may go beyond a better rate-distortion performance but also improve the meaningfulness of the motion estimates. This fact may be important for next generation coders that build models of the scene from the low-level information used in the current standards.

In this paper, a novel subjective redundancy removal procedure [13], [14] and a novel perceptually weighted motion estimation algorithm [12] are jointly considered to present a fully perceptual motion compensated video coder. The aim of the

Manuscript received May 6, 1999; revised June 18, 2001. This work was supported in part by CICYT Projects TIC 1FD97-0279 and TIC 1FD97-1910. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rashid Ansari.

J. Malo and J. M. Artigas are with the Departament d'Òptica, Universitat de València, 46100 Burjassot, València, Spain (e-mail: Jesus.Malo@uv.es; http://taz.uv.es/~jmalo).

J. Gutiérrez, I. Epifanio, and F. J. Ferri are with the Departament d'Informàtica Universitat de València, 46100 Burjassot, València, Spain.

Publisher Item Identifier S 1057-7149(01)08211-2.

paper is to assess the relative relevance of optimal variable-size BMAs and quantizer improvements. To this end, the decoded frames are explicitly compared and analyzed in terms of perceptually meaningful distortion measures [15], [16]. The meaningfulness of the motion information is tested by using it as input for a well established motion-based segmentation algorithm used in model-based video coding [17], [18].

The paper is organized as follows. In Section II, the current methods for quantizer design and variable-size BMA for motion compensation are briefly reviewed. The proposed improvements in the quantizer design, along with their perceptual foundations, are detailed in Section III. In Section IV, the proposed motion refinement criterion is obtained from the requirement of a monotonic reduction of the significant (perceptual) entropy of DFD and DVF. The comparison experiments are presented and discussed in Section V. Some final remarks are given in Section VI.

II. Conventional Techniques for Transform Quantizer Design and Multigrid Motion Estimation

The basic elements of a motion compensated coder are the optical flow estimation and the prediction error quantization. The optical flow information is used to reduce the objective temporal redundancy, while the quantization of the transformed error signal [usually a 2-D discrete cosine transform (DCT)] reduces the remaining (objective and subjective) redundancy to certain extent [1]–[4].

Signal independent JPEG-like uniform quantizers are employed in the commonly used standards [1]–[4]. In this case, bit allocation in the 2-D DCT domain is heuristically based on the threshold detection properties of the HVS [2], [3], but neither amplitude nonlinearities [19] nor temporal properties of the HVS [20]–[22] are taken into account. The effect of these properties is not negligible [23], [24]. In particular, the nonlinearities of the HVS may have significant effects on bit allocation and improve the subjective results of the JPEG-like quantizers [13], [14], [25], [26].

The conventional design of a generic transform quantizer is based on the minimization of the *average* quantization error over a training set [27]. However, the techniques based on average error minimization have some subjective drawbacks in image coding applications. The optimal quantizers (in an average error sense) may underperform on individual blocks or frames [9] even if the error measure is perceptually weighted [28]: the accumulation of quantization levels in certain regions in order to minimize the average perceptual error does not ensure good behavior on a particular block of the DFD. This suggests that the subjective problems of the conventional approach are not only due to the use of perceptually unsuitable metrics, as usually claimed, but are also due to the use of an inappropriate *average error* criterion. In addition to this, quantizer designs that depend on the statistics of the input have to be re-computed as the input signal changes. These factors favor the use of quantizers based on the threshold frequency response of the HVS instead of the conventional, average error-based quantizers.

Multigrid motion estimation techniques are based on matching between variable-size blocks of consecutive frames of the sequence [6]. The motion estimation starts at a coarse resolution (large blocks). At a given resolution, the best displacement for each block is computed. The resolution of the motion estimate is locally increased (a block of the quadtree is split) according to some refinement criterion. The process ends when no block of the quadtree can be split further.

The splitting criterion is the most important part of the algorithm because it controls the local refinement of the motion estimate. The splitting criterion has effects on the relative volumes of DVF and DFD [7]–[12], and may give rise to unstable motion estimates due to an excesive refinement of the quadtree structure [12], [29]. The usefulness of the motion information for higher-level purposes (as in model-based video coding [5], [30], [31]) highly depends on its robustness (absence of false alarms) and hence on the splitting criterion. Motion-based segmentation algorithms [17], [18] require reliable initial motion information, especially when using sparse (nondense) flows such as those given by variable-size BMA. Two kinds of splitting criteria have already been used: 1) the magnitude of the prediction error, e.g., energy, mean-square error or mean-absolute error [6], [29], [32], [33], and 2) the complexity of the prediction error. In this case, the zeroth-order spatial entropy [7], [8] and the entropy of the encoded DFD [9]–[12] have been reported. While the magnitude-based criteria were proposed without a specific relation to the encoding of the DFD, the entropy-based criteria make explicit use of the trade off between DVF and DFD.

Since the first entropy-constrained approach was introduced [7], [8], great effort has been devoted to obtaining analytical [9]–[11] or numerical [12] optimal entropy-constrained quadtree DVF decompositions. These approaches criticize the (faster) entropy measure of the DFD in the spatial domain of Dufaux *et al.* because it does not take into account the effect of the selective DCT quantizer. This necessarily implies a suboptimal bit allocation between DVF and DFD. The literature [9]–[12] reports the optimality of the proposed methods, but the practical (subjective) effect of this gain on the reconstructed sequence is not analyzed. In particular, only perceptually unweighted SNR or MSE distortion measures are given and no explicit comparison of the decoded sequences is shown.

III. Perceptually Uniform DCT Quantization

Splitting the original signal into two lower complexity signals (DVF and DFD) does reduce their redundancy to a certain extent. However, the enabling fact behind very-low-bit-rate coding is that not all the remaining data are significant to the human observer. This is why more than just the strictly predictable data can be safely discarded in the DFD quantization.

According to the current models of human contrast processing and discrimination [34], [35], the input spatial patterns are first mapped onto a local frequency domain through a set of bandpass filters with different relative gains. After that, a log-like nonlinearity is applied to each transform coefficient to obtain the response representation. Let us describe this two-step process as

$$\mathbf{A} \xrightarrow{T} \mathbf{a} \xrightarrow{R} \mathbf{r} \tag{1}$$

where

vector $\mathbf{A}$	input image;
matrix T	filter bank;
vector $\mathbf{a}$	local frequency transform;
function R	nonlinearity;
vector $\mathbf{r}$	response to the input.

The n components of the image vector, $\mathbf{A} = \{A_x\}_{x=1}^{n}$, represent the samples of the input luminance at the discrete positions $x = 1, \ldots, n$. T is a $m \times n$ matrix constituted by the impulse responses of the m bandpass filters. The local frequency transform is $\mathbf{a} = T \cdot \mathbf{A}$. Each coefficient of the transform $\mathbf{a} = \{a_f\}_{f=1}^{m}$, represent the output of the filter f with $f = 1, \ldots, m$. Each local filter f is tuned to a certain frequency. In general [34], each coefficient r_f of the response $\mathbf{r} = R(\mathbf{a}) = \{r_f\}_{f=1}^{m}$ will depend on several transform coefficients $a_{f'}$. However, at a first approximation [19], the contributions of $a_{f'}$ with $f' \neq f$ can be neglected.

The effect of the response R in the transform $\mathbf{a}$ can be conveniently modeled by a nonuniform perceptual quantizer Q_p. This interpretation as a quantizer is based on the limited resolution of the HVS. If the amplitude of a basis function of the transform T is modified, the induced perception will remain constant until the just noticeable difference (JND) is reached. In this case, as in quantization, a continuous range of amplitudes gives rise to a single perception [36], [37]. This perceptual quantizer has to be nonuniform because the empirical JNDs are nonuniform [19], [21], [22], [34]. The similarity between the impulse responses of the perceptual filters of the transform T and the basis functions of the local frequency transforms used in image and video coding has been used to apply the experimental properties of the perceptual transform domain to the block DCT transform as a reasonable approximation [13], [14], [25], [26], [38], [39]. In this paper, Q_p is formulated in the DCT domain through an explicit design criterion based on a distortion metric that includes the HVS nonlinearities [15], [16] and some temporal perceptual features [20]–[22].

A. Maximum Perceptual Error (MPE) Criterion for Quantizer Design

The natural way of assessing the quality of an encoded picture (or sequence) involves a one-to-one comparison between the original and the encoded version. The result of this comparison is related to the ability of the observer to notice the particular quantization noise in the presence of the original (masking) pattern. This one-to-one noise detection or assessment is clearly related to the tasks behind the standard pattern discrimination models [34], [35], in which an observer has to evaluate the distortion from a masking stimulus. In contrast, a hypothetical request of assessing the global performance of a quantizer over a set of images or sequences would involve a sort of averaging of each one-to-one comparison. It is unclear how a human observer does this kind of averaging to obtain a global feeling of performance and the task itself is far from the natural one-to-one comparison that arises when one looks at a particular picture.

The conventional techniques of transform quantizer design use average design criteria in such a way that the final quantizer achieves the minimum average error over the training set (sum of the one-to-one distortions weighted by their probability) [27]. However, the minimization of an average error measure does not guarantee a satisfactory subjective performance on individual comparisons [9]. Even if a perceptual weighting is used, the average criteria may bias the results. For instance, Macq [28] used uniform quantizers instead of the optimal Lloyd-Max quantizers [27], [40], due to the perceptual artifacts caused by the outliers on individual images.

To prevent large perceptual distortions on individual images arising from outlier coefficients, the coder should restrict the *maximum perceptual error (MPE)* in each coefficient and amplitude [13], [14]. This requirement is satisfied by a perceptually uniform distribution of the available quantization levels in the transform domain. If the perceptual distance between levels is constant, the MPE in each component is bounded regardless of the amplitude of the input.

In this paper, the restriction of the MPE will be used as a design criterion. This criterion can be seen as a perceptual version of the minimum maximum error criterion [9]. This idea has been implicitly used in still image compression [25], [26] to achieve a constant error contribution from each frequency component on an individual image. It has been shown that bounding the perceptual distortion in each DCT coefficient may be subjectively more effective than minimizing the average perceptual error [13], [14]. Moreover, the MPE quantizers reduce to the JPEG and MPEG quantizers if a simple (linear) perception model is considered.

B. Optimal Spatial Quantizers Under the MPE Criterion

The design of a transform quantizer for a given block transform involves finding the optimal number of quantization levels for each coefficient (bit allocation) and the optimal distribution of these quantization levels in each case [27].

Let us assume that the squared perceptual distance between two similar patterns in the transform domain $\mathbf{a}$ and $\mathbf{a} + \Delta\mathbf{a}$ is given by a weigthed sum of the distortion in each coefficient

$$D^2(\mathbf{a}, \mathbf{a} + \Delta\mathbf{a}) = \sum_{f=1}^{m} D_f^2 = \sum_{f=1}^{m} W_f(a_f)\, \Delta a_f^2 \qquad (2)$$

where $W_f(a_f)$ is a frequency and amplitude-dependent perceptual metric.

In order to prevent large perceptual errors on individual images coming from outlier coefficient values, the coder should be designed to bound the MPE for every frequency f and amplitude a_f.

If a given coefficient (at frequency f) is represented by N_f quantization levels distributed according to a density $\lambda_f(a_f)$ the maximum Euclidean quantization error at an amplitude a_f will be bounded by half the Euclidean distance between two levels

$$\Delta a_f(a_f) \leq \frac{1}{2N_f\, \lambda_f(a_f)}. \qquad (3)$$

The MPE for that frequency and amplitude will be related to the metric and the density of levels:

$$\mathrm{MPE}_f(a_f) = W_f(a_f) \cdot \max\left(\Delta a_f(a_f)\right)^2 = \frac{W_f(a_f)}{4N_f^2\, \lambda_f^2(a_f)}. \qquad (4)$$

The only density of quantization levels that gives a constant MPE bound over the amplitude range is the one that varies as the square root of the metric

$$\lambda_{f_{\mathrm{opt}}}(a_f) = \frac{W_f(a_f)^{1/2}}{\int W_f(a_f)^{1/2} da_f}. \tag{5}$$

With these optimal densities, the MPE in each coefficient f will depend on the number of allocated levels and on the integrated value of the metric

$$\mathrm{MPE}_{f\mathrm{opt}} = \frac{1}{4N_f^2}\left(\int W_f(a_f)^{1/2} da_f\right)^2. \tag{6}$$

Fixing the same maximum distortion for each coefficient $\mathrm{MPE}_{f\mathrm{opt}} = k^2$ and solving for N_f, the optimal number of quantization levels is obtained

$$N_{f_{\mathrm{opt}}} = \frac{1}{2k}\int W_f(a_f)^{1/2} da_f. \tag{7}$$

The general form of the optimal MPE quantizer is given by (5) and (7) as a function of the perceptual metric. Thus, the behavior of the MPE quantizer will depend on the accuracy of the selected W. Here, a perceptual metric related to the gradient of the nonlinear response R and to the amplitude JNDs has been considered [15], [16]

$$\begin{aligned} W_f(a_f) &= \left(\frac{\partial R(\mathbf{a})}{\partial a_f}\right)^2 \propto \mathrm{JND}(a_f)^{-2} \\ &= \left(\mathrm{CSF}_f^{-1} + \frac{a_f}{L} G_f(a_f)\right)^{-2} \end{aligned} \tag{8}$$

where L is the local mean luminance, CSF_f is the *contrast sensitivity function* (the bandpass linear filter which characterizes the HVS performance for low amplitudes [20], [24], [28], [41]), and $G_f(a_f)$ are empirical monotonically increasing functions of amplitude for each spatial frequency to fit the amplitude JND data [15]. In particular, we have used the CSF of Nygan *et al.* [41]

$$\begin{aligned} \mathrm{CSF}_f = &\left(\frac{1}{4} + \frac{1}{\pi^2}\log^2\left(\frac{2\pi f}{11.6} + \sqrt{\frac{4\pi^2 f^2}{11.6^2} + 1}\right)\right)^{\frac{1}{2}} \\ &\times (115.9 + 258.1 f)\cdot e^{-0.29f} \end{aligned} \tag{9}$$

and the following nonlinear functions $G_f(a_f)$ [15] (frequency f in cycles/degrees)

$$G_f(a_f) = \frac{(-0.03\log f + 0.3)\left(\frac{a_f}{L}\right)^{\frac{0.8 f^{1.7}}{0.5 + f^{1.7}}} - \mathrm{CSF}_f^{-1}}{\left((-0.03\log f + 0.3)\mathrm{CSF}_f\right)^{-\frac{0.5 + f^{1.7}}{0.8 f^{1.7}}} + \frac{a_f}{L}}. \tag{10}$$

It is important to note that the metric weight for each coefficient in (8) has two contributions: one constant term (the CSF) and one amplitude-dependent term that vanishes for low amplitudes. This second term comes from a nonlinear correction to the linear threshold response described by the CSF. These two terms in the metric give two interesting particular cases of the MPE formulation. First, if a simple linear perception model is assumed, a CSF-based MPEG-like quantizer is obtained. If the nonlinear correction in (8) is neglected, uniform quantizers are obtained for each coefficient and N_f becomes proportional to the CSF, which is one of the recommended options in the JPEG and MPEG standards [1]–[3]. Second, if both factors of the metric are taken into account, the algorithm of [13], [14], [26] is obtained: the quantization step size is input-dependent and proportional to the JNDs and bit allocation is proportional to the integral of the inverse of the JNDs. From now on, these two cases will referred to as linear and nonlinear MPE, respectively.

The CSF-based (linear MPE) quantizer used in MPEG [1]–[3] and the proposed nonlinear MPE quantizer [13], [14], [26], represent different degrees of approximation to the actual quantization process, Q_p, eventually carried out by the HVS. The scheme that takes into account the perceptual amplitude nonlinearities will presumably be more efficient in removing the subjective redundancy from the DFD.

Fig. 1 shows the product $N_f \cdot \lambda_f(a_f)$ for the linear (MPEG-like) and the nonlinear MPE quantizers. This product represents the number of quantization levels per unit of area in the frequency and amplitude plane. This surface is a useful description of where a quantizer concentrates the encoding effort [14]. Fig. 2 shows the bit allocation solutions (number of quantization levels per coefficient N_f) in the linear and the nonlinear MPE cases. Note how the amplitude nonlinearities enlarge the bandwidth of the quantizer in comparison to the CSF-based case. This enlargement will make a difference when dealing with wide spectrum signals like the DFD.

C. Introducing HVS Temporal Properties in the Prediction Loop

The previous considerations about optimal MPE 2-D transform quantizers can be extended to three-dimensional (3-D) spatio-temporal transforms. The HVS motion perception models extend the 2-D spatial filter bank to nonzero temporal frequencies [42], [43]. The CSF filter is also defined for moving gratings [20] and the contrast discrimination curves for spatio-temporal gratings show roughly the same shape as the curves for still stimuli [21], [22]. By using the 3-D CSF and similar nonlinear corrections for high amplitudes, the expression of (8) could be employed to measure differences between local moving patterns. In this way, optimal MPE quantizers could be defined in a spatio-temporal frequency transform domain. However, the frame-by-frame nature of any motion compensated scheme makes the implementation of a 3-D transform quantizer in the prediction loop more difficult.

In order to exploit the subjective temporal redundancy removal to some extent, the proposed 2-D MPE quantizer can be complemented with one-dimensional (1-D) temporal filtering based on the perceptual bit allocation in the temporal dimension. This temporal filter can be implemented by a simple finite impulse response weighting of the incoming error frames. The temporal frequency response of the proposed 1-D filter is set proportional to the number of quantization levels that should be allocated in each temporal frequency frame of a 3-D MPE optimal quantizer. For each spatio-temporal coefficient $\mathbf{f} = (f_x, f_t)$, the optimal number of quantization levels is given

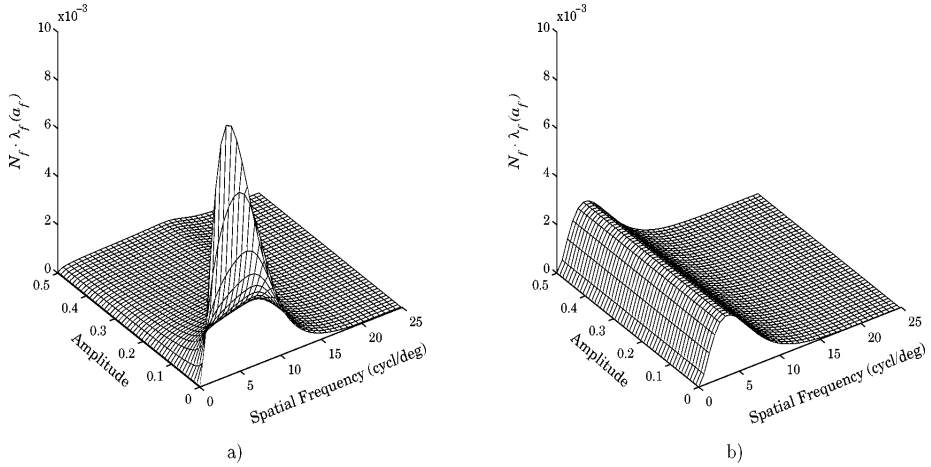


Fig. 1. Relative number of quantization levels allocated in the frequency and amplitude plane for a) nonlinear MPE and b) linear MPE quantizers. The surfaces are scaled to have unit integral (the same total number of quantization levels). The distribution of the quantization levels in amplitude for a certain coefficient is just the corresponding slice of the surface at the desired frequency. The MPE design [(5) and (7)] implies that this surface is proportional to the metric $N_f \cdot \lambda_f(a_f) \propto W(a_f)^{1/2}$ so different perception models (different metrics) give rise to a different distribution of quantization levels. Note that the distribution is uniform for every frequency in the linear MPE (MPEG-like) case and nonuniform (peaked at low amplitides) in the nonlinear MPE case.

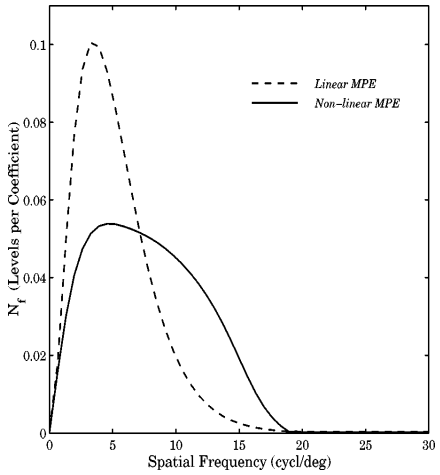


Fig. 2. Bit allocation results (relative number of quantization levels per coefficient) for the linear MPE (MPEG-like case) and for the nonlinear MPE case. The curves are scaled to have unit integral (the same total number of quantization levels). In the linear case, the metric is just the square of the CSF and then $N_f \propto \mathrm{CSF}_f$ as recommended by JPEG and MPEG. A more complex (nonlinear) model gives rise to a wider quantizer bandpass.

by (7). Integrating over the spatial frequency, the number of quantization levels for that temporal frequency is

$$N_{f_t} = \sum_{f_x} N_{\mathbf{f}} = \frac{1}{2k} \sum_{f_x} \int W_{\mathbf{f}}(a_{\mathbf{f}})^{1/2}\, da_{\mathbf{f}}. \qquad (11)$$

Fig. 3(a) shows the number of quantization levels for each spatio-temporal frequency of a 3-D nonlinear MPE quantizer. This is the 3-D version of the 2-D nonlinear bit allocation of Fig. 2. Note that (except for a scale factor) the spatial frequency curve for the zero temporal frequency is just the solid curve of Fig. 2. Fig. 3(b) shows the temporal frequency response that is obtained by integrating over the spatial frequencies.

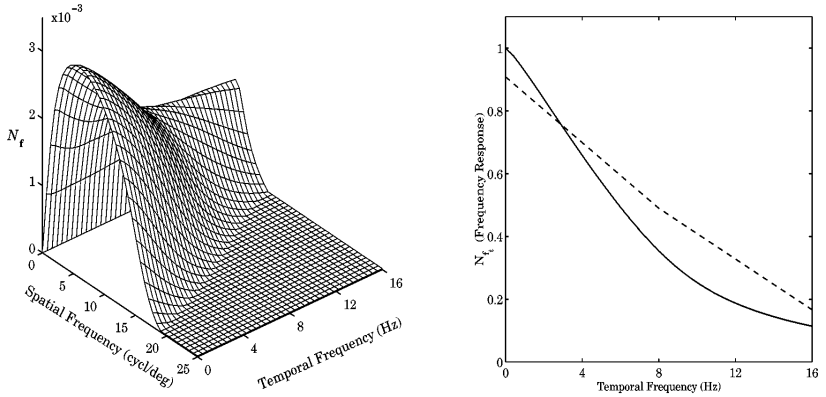
IV. Perceptual Feedback in the Motion Estimation

Any approximation to the actual perceptual quantization process Q_p has an obvious application in the DFD quantizer design, but it may also have interesting effects on the computation of the DVF if the proper feedback from the DFD quantization is established in the prediction loop.

If all the details of the DFD are considered to be of equal importance, we would have an *unweighted* splitting criterion as in the difference-based criteria [6], [29], [32], [33] or as in the spatial entropy-based criterion of Dufaux *et al.* [7], [8]. However, as the DFD is going to be simplified by some nontrivial quantizer Q_p, which represents the selective bottleneck of early perception, not every additional detail predicted by a better motion compensation will be significant to the quantizer. In this way, the motion estimation effort has to be focused on the moving regions that contain *perceptually significant motion information*. In order to formalize the concept of perceptually significant motion information, the work of Watson [36] and Daugman [37] on entropy reduction in the HVS should be taken into account. They assume a model of early contrast processing based on a pair (T, Q_p) and suggest that the entropy of the cortical scene representation (a measure of the perceptual entropy of the signal) is just the entropy of the quantized version of the transformed image. Therefore, a measure of the perceptual entropy H_p of a signal $\mathbf{A}$ is

$$H_p(\mathbf{A}) = H\left(Q_p[\mathrm{T} \cdot \mathbf{A}]\right). \qquad (12)$$

Using this perceptual entropy measure (which is simply the entropy of the output of a MPE quantizer), we can propose an explicit definition of what perceptually significant motion information is. Let us motivate the definition as follows. Given a cer-



a) b)

Fig. 3. (a) Nonlinear MPE bit allocation results in the 3-D spatio-temporal frequency domain (relative number of quantization levels per 3-D coefficient). The surface is scaled to have unit integral. (b) Frequency response of the perceptual temporal filter, proportional to N_{ft}. The solid line corresponds to the theoretical curve and the dashed line stands for the actual frequency response obtained with the fourth-order FIR filter used in the experiments (see Section V). The coefficients of the filter in the temporal domain are: 0.0438, 0.1885, 0.4443, 0.1885, and 0.0438.

tain motion description, DVF, with entropy $H(\text{DVF})$, a prediction of the next frame can be done. Some particular error, DFD, will be obtained, with a perceptual entropy $H_p(DFD)$. If additional motion information is available, $\Delta H(\text{DVF}) > 0$ (more complex quadtree segmentation and more motion vectors), one would expect a reduction of the perceptual information of the remaining DFD, i.e., $\Delta H_p(DFD) < 0$. Let us define this additional motion information as perceptually significant only if it implies a greater reduction in the perceptual entropy of the prediction errors: $\Delta(DVF)$ *is perceptually significant if*

$$\Delta H(DVF) < -\Delta H_p(DFD). \tag{13}$$

Broadly speaking, some additional motion information is perceptually significant if it increments the perceptual information of the prediction more than its own volume.

If each motion refinement in a variable-size BMA, $\Delta H(\text{DVF}) = H(\text{DVF}_{\text{split}}) - H(\text{DVF}_{\text{nosplit}})$, is required to be perceptually significant, the following *perceptually weighted* splitting criterion arises: *a block of the quadtree structure should be split if*

$$H(\text{DVF}_{\text{split}}) + H_p(\text{DFD}_{\text{split}}) < H(\text{DVF}_{\text{nosplit}}) + H_p(\text{DFD}_{\text{nosplit}}) \tag{14}$$

where $H(\text{DVF})$ is the entropy of the DPCM coded DVF plus the information needed to encode the quadtree structure and $H_p(\text{DFD})$ is the perceptual entropy of the residual error signal.

Equation (14) has the same form as the criterion proposed by Dufaux and Moscheni [7], [8], except for the way in which the entropy of the DFD is computed. In this case, the unweighted entropy of the DFD in the spatial domain is replaced by the perceptually weighted entropy measure in the appropriate encoding domain. The consideration of the entropy of the quantized DFD (or perceptual entropy) is the main difference between the suboptimal approach of Dufaux and Moscheni [7], [8], and the optimal approaches [9]–[12]. The optimal approaches only differ in the way the problem is stated: While in [9]–[11] an explicit rate-distortion sum is minimized, in [12] a fixed distortion is assumed (constant MPE quantizer) and a monotonically decreasing behavior for the rate is imposed. These optimal algorithms do not necessarily find the absolute minimum in the rate-distortion sense but only local minima; however, as the actual DFD entropy is used, the performance of the suboptimal result is always improved. It is interesting to note that the reasoning about the perceptual relevance of the motion information presented here leads to (14), which takes into account the quantized DFD entropy, in a natural way. As long as [9]–[12] use CSF-based quantizers, all these optimal approaches can also be referred to as perceptually weighted variable-size BMAs. In what follows, the algorithm of Dufaux *et al.* [7], [8] and the algorithm from (14) will be compared and referred to as unweighted and perceptually weighted variable-size BMA, respectively.

The perceptual quantizer constraint on the motion estimate comes from the particular video coding application in which the actual bit-rate has to be minimized [9]–[12]. However, the benefits of including the properties of the biological filters in the motion estimation may go beyond the rate-distortion optimization. The bandpass shape of human sensitivity (Fig. 2) gives a scale-dependent measure of the perceptual entropy of the DFD. As some frequency bands (some scales) have more perceptual importance than others, the application of the perceptual criterion results in a different splitting behavior in the different levels of the multigrid structure. Fig. 4 qualitatively shows how a bandpass criterion may give a scale-dependent splitting result. In coarse levels of the multigrid (left side figures), the spatial support of the DFD is large due to the displacement of large blocks. The uncertainty relationship $\Delta x \cdot \Delta f = k$ leads to a DFD with narrow bandwidth in the case of a large DFD support. Conversely, in the fine levels of the multigrid (right-side

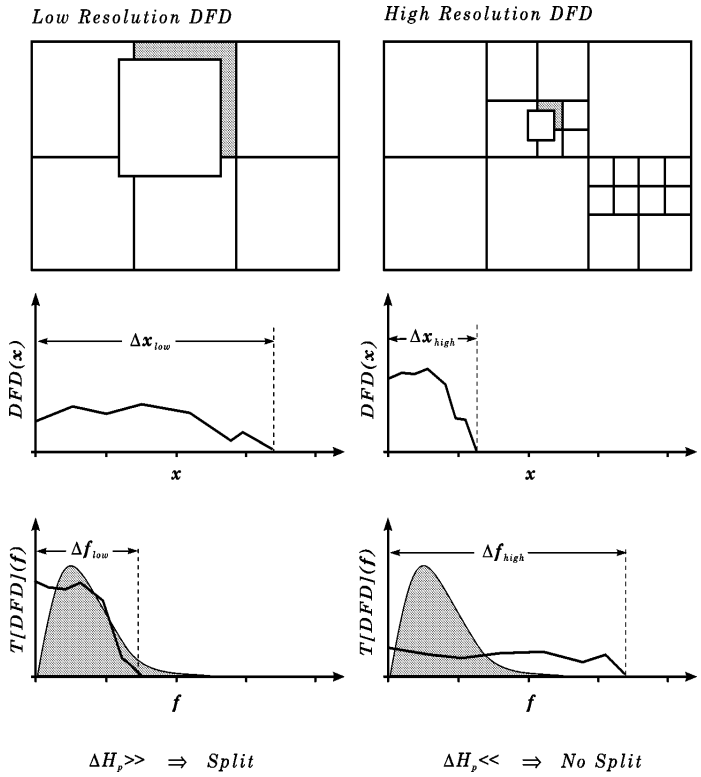


Fig. 4. Scale-dependent splitting strategy due to perceptual feedback. The dashed regions in the frequency domain represent the bit allocation of the MPE quantizer. They determine the frequency band which is considered to compute the perceptual entropy of the signal. For a given energy and resolution level, the spatial extent and the frequency bandwidth of the DFD (thick solid lines) are related by the uncertainty relation $\Delta x \cdot \Delta f = k$. The bandwidth of the DFD will depend on the resolution, giving rise to a different splitting behavior when using a bandpass splitting criterion such as the perceptual entropy.

figures), the DFD is spatially localized, giving rise to a broadband error signal. If the complexity measure is more sensitive to the complexity of the signal in low- and middle-frequency bands, the splitting criterion will be tolerant in the coarse levels of the multigrid and will be strict in the high-resolution levels. The next section will show that this scale-dependent behavior is useful for discriminating between significant and noisy motions.

V. Experiments and Discussion

Four experiments on several standard sequences [44] were carried out at a fixed bit-rate:

- Different quantizers with the same motion estimation.
- Different motion estimations with the same quantizer.
- Relative relevance of the improvements in the motion estimation and the quantization.
- Proposed scheme versus previous comparable schemes (H.263, MPEG-1).

We are interested in two different comparisons: 1) quality of the reconstructed signal and 2) usefulness of the motion flows for higher-level purposes. To this end, examples of the reconstructed frames and subjective distortion measures using a perceptually meaningful metric [15], [16] are given in each case. Also, a well-known motion-based segmentation algorithm for high-level video coding [17], [18] has been used with both perceptually weighted and perceptually unweighted BMAs to test their respective usefulness.

In every experiment, the quantizer was adapted for each group of pictures to achieve the desired bit-rate (200 kb/s with QCIF format). In order to highlight the relative differences among the different approaches considered, only the first frame was intracoded and only forward prediction was used in the remaining frames (i.e., no bidirectional interpolated frames nor additional intracoded frames were introduced). As a consequence, the results may seem abnormally distorted at this rate. The DC coefficient of error signals is basically zero [45], so the luminance information of the original blocks was used to normalize the amplitudes of the error blocks. These luminance values were not quantized but DPCM coded from block to block as in the JPEG standard. The 1-D temporal filter was implemented by a linear-phase FIR filter using least-squares error minimization in the frequency response. A simple fourth-order filter was used to restrict the buffer requirements (see Fig. 3 for the achieved frequency response and the filter coefficients).

A maximum of five resolution levels (blocks from $64 \times 64$to 4×4) were used in the variable-size BMA quadtrees. Blocks of size 8×8 were used in the fixed-size BMA. The *n*-step displace-

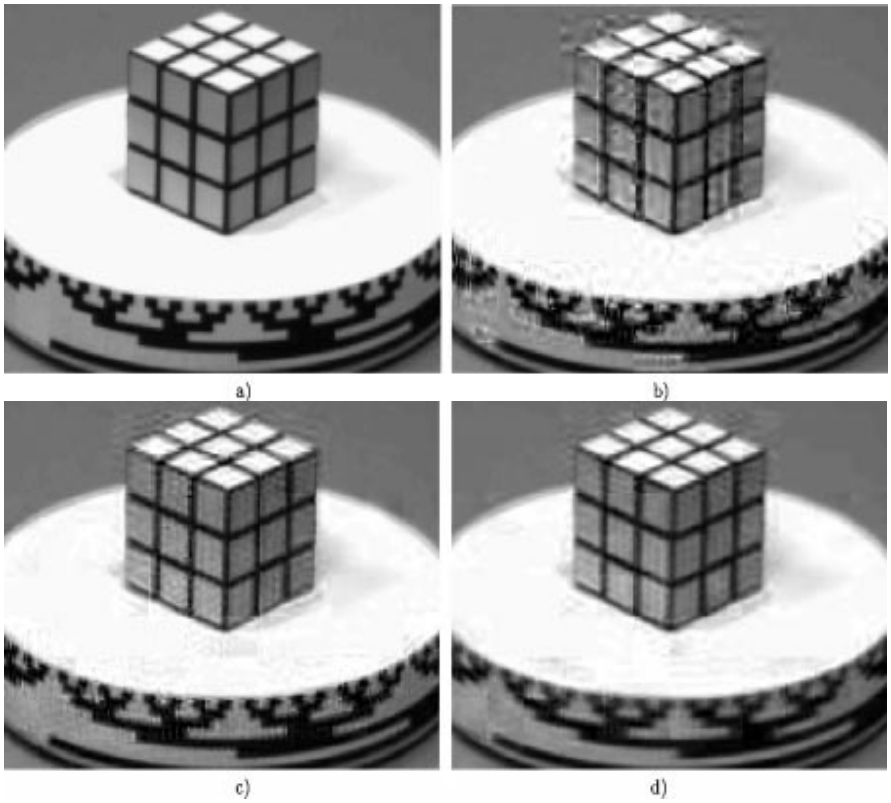


Fig. 5. Quantization results with a fixed motion estimation algorithm (unweighted variable-size BMA). (a) Original (detail of frame 7 of the *Rubik* sequence). (b) Two-dimensional linear MPE, uniform MPEG-like quantization. (c) Two-dimensional nonlinear MPE. (d) Two-dimensional nonlinear MPE and temporal filtering.

ment search [8] and integer-pixel accuracy was used in every resolution level of the BMAs. The usual correlation was used as a similarity measure.

The definitive proof of the subjective benefits of an algorithm is a set of psychophysical experiments on the quality of the reconstructed sequences, but it requires time-consuming experiences involving several observers. An approximate, but more practical, approach is to use perceptually meaningful distortion measures. This is not a simple issue [46], [47]. However, basic facts as the spatial frequency sensitivity make a fundamental difference between plain Euclidean distortion metrics (e.g MSE or SNR) and any (even the simplest) perceptually weighted metric. In the experiments, the perceptual distortions were computed through (2) using the nonlinear perceptual metric of (8) [15], [16] on a frame-by-frame basis. This metric incorporates the basic elements of early achromatic visual processing [46], [47]: luminance adaptation, spatial frequency channels, frequency dependent filtering, contrast masking, and (quadratic) probability summation. The squared distortion was computed for each DCT block in a frame and averaged across the frame. This straighforward frame-by-frame implementation may neglect some temporal factors, but it still gives a rough approximation to the observers opinion and is thus useful for confirming the results.

A. Experiment 1: Different Quantizers with the Same Motion Estimation

The performance of the linear MPE (MPEG-like) quantizer and the proposed nonlinear MPE quantizers was compared using the same (perceptually unweighted H.263-like) motion estimation. Fig. 5 shows a representative example of the kind of errors obtained with the different quantizers at a fixed bit-rate. Fig. 6 shows the increase of the perceptual distortion in the different reconstructions of the *Rubik* sequence.

The consideration of the (2-D or 3-D) nonlinearities introduces a substantial improvement in comparison to the linear MPE quantizer. The temporal filtering smooths the reconstructed sequence and reduces to some extent the remaining blocking effect and busy artifacts of the 2-D nonlinear approach. However, despite the eventual visual advantages of this temporal filtering, the key factor in the improvements of the proposed quantizers is the consideration of the amplitude nonlinearities and the corresponding enlargement of the spatial quantizer bandwidth (Fig. 2). The reconstructed examples and the computed distortion confirm this point. This enlargement implies that the quantized signal keeps some significant details otherwise discarded, avoiding the rapid degradation of the reconstructed signal.

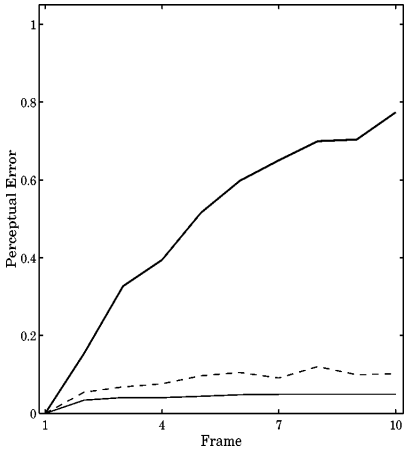


Fig. 6. Perceptual distortion measures for the frames of the *Rubik* sequence using different quantizers with the same (unweighted H.263-like) motion estimation. All the distortion results presented in the paper are normalized by the worst distortion obtained at the last frame of the *Rubik* sequence using an H.261 or MPEG-1 approach. The thick solid line correspond to the linear MPE quantizer. The thin solid line correspond to the 2-D nonlinear MPE quantizer and the thin dashed line correspond to the 2-D nonlinear quantizer with temporal filtering. In this 3-D case, the frame-by-frame implementation of the perceptual distortion measure may slightly overestimate the visual effect of the frame blurring introduced by the temporal filter. Despite this fact, the perceptual distortions reveal that the amplitude nonlinearities substantially improve the reconstruction results.

A very interesting consequence is that the (2-D or 3-D) nonlinear quantizers keep the distortion bounded over a large group of frames (compare the behavior of the linear and nonlinear distortion curves). In this way, the need to introduce bit-consuming intracoded frames is reduced.

B. Experiment 2: Different Motion Estimations with the Same Quantizer

The proposed perceptually weighted variable-size BMA and the unweighted variable-size BMA were compared in this experiment using the same linear MPE (MPEG-like) quantizer. The results with a fixed-size 8×8 BMA are also included as a reference.

The quantitative rationale behind entropy constrained approaches is saving motion information to improve the DFD encoding and have a better reconstruction. From this point of view, the only reason to take into account the perceptual quantization is an improvement of the reconstructed sequence. The qualitative rationale to take into account the perceptual quantization is to include a perceptual criterion to decide when some motion information is significant. Accordingly, the effects of the perceptual weight in the motion estimation were tested in two different ways. First, the effect of the savings in motion information on the signal quality was analyzed. Second, the usefulness of the motion information for higher-level purposes was tested.

The Table I and Fig. 7 analyze the bit allocation performance of the motion estimation algorithms. Table I shows the percentage of the total bit-rate used for the motion flow. Fig. 7 shows the reduction of the bit-rate in the *Taxi* sequence while refining the motion estimate with the weighted and unweighted variable-size BMAs. The consideration of the perceptual quantizer in the motion estimation certainly minimizes the entropy of the motion flow and the total entropy as claimed in [9]–[12]. The problem with the unweighted criterion is that it is too permissive at high-resolution levels. In this way, too many blocks are split increasing the motion information without decreasing the information content of the DFD, i.e., perceptually negligible motion information is being added.

However, do these motion information savings in [9]–[12] have a significant effect on the reconstructed quality? Fig. 8 shows a representative example of the decoded results using the same MPEG-like quantizer and the different considered motion estimations at a fixed bit-rate (frame 7 of the *Taxi* sequence). Fig. 9 shows the increase of the perceptual distortion in the different reconstructions of the *Taxi* sequence. The distortions here are relatively lower than in the *Rubik* sequence because the area of the moving regions (and hence the distorted area) is also lower. These results show that both variable-size BMAs make some difference in quality with regard to the fixed-size H.261 BMA due to the savings in DVF information. However, the practical advantages of the better bit allocation between DVF and DFD over the suboptimal variable-size BMA are not evident. The benefits of the better bit allocation in comparison to the suboptimal variable-size BMA are so small (Table I and Fig. 7) that in practice (Figs. 8 and 9), they cannot be exploited by the quantizer to give a better encoded DFD. That is, any adaptive DVF consumes such a small portion of the total bit-rate that there is no significant difference in the reconstruction between the different variable-size BMA algorithms.

The scale-independent behavior of the unweighted algorithm leads to too many splittings at high resolution. This suboptimal result has no effect in the reconstructed sequence, but it may give rise to noisy, less meaningful, motion information. In order to test the meaningfulness of the motion information obtained with the different algorithms, a well-known motion-based segmentation algorithm [17], [18] was initialized with the different flows.

The layer identification algorithm [17], [18] starts from an arbitrary initial segmentation, which is refined by estimating the affine motion model of each region and merging the regions with similar affine parameters. Given a pair of frames from a sequence with n moving objects, the segmentation algorithm gives m masks, representing the identified objects and the corresponding affine models.

In order to assess the relative performance of the flows under consideration, the following segmentation error measure has been used. Let ϵ_i be the dissimilarity between each (manually segmented) object O_i and its corresponding mask M (M is the mask that maximizes $M_j \cap O_i$ for all the obtained masks M_j)

$$\epsilon_i = \int_{M \cap O_i^C} D_i(x)\, dx + \int_{O_i \cap M^C} D_i(x)\, dx \qquad (15)$$

where C stands for the set complement and $D_i(x)$ is the Chamfer distance, i.e., the distance from x to the boundary of the object O_i [48]. This distance progressively penalizes unmatched pixels that are far from the object boundary. The

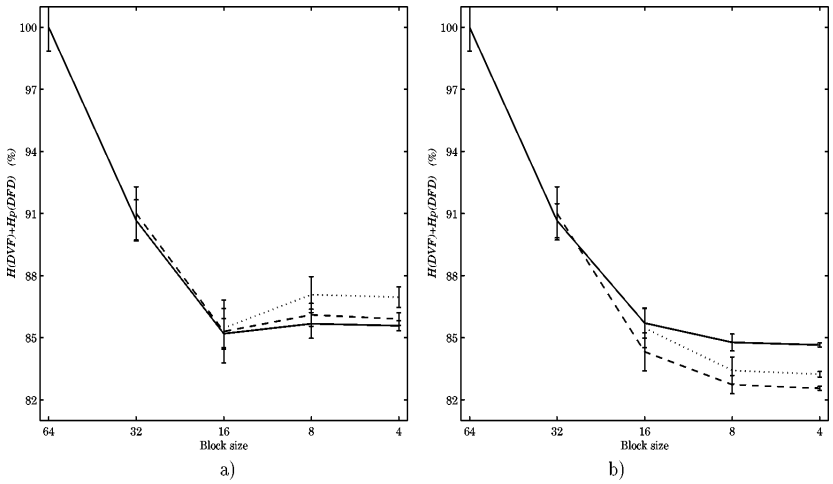


Fig. 7. Bit-rate of DVF and DFD while refining the motion estimate. The values are given as a percentage of the total entropy at the lowest resolution level. The curves represent the average percentage of the bit-rate across the frames and the error bars represent the standard deviation. The different lines indicate different starting resolutions (initial block sizes): solid 64×64, dashed 32×32, and dotted 16×16. (a) Unweighted spatial entropy splitting criterion. (b) Perceptually weighted splitting criterion.

TABLE I
PERCENTAGE OF THE TOTAL BIT-RATE USED FOR THE MOTION FLOW (DVF)

	TAXI	RUBIK	YOSEMITE	TREES	Average
FSBMA	12.57	21.71	46.19	39.21	30 ± 8
Unweight. VSBMA	1.89	3.29	8.69	4.97	4.7 ± 1.5
Weighted VSBMA	0.94	1.87	4.32	3.24	2.6 ± 0.7

segmentation error for a frame ϵ is the sum of the individual segmentation errors ϵ_i. It has been empirically found that this error measure adequately describes the intuitive quality of the segmentation (Figs. 10 and 11 and their particular errors illustrate this point).

Several experiments were carried out using the layer identification algorithm [17], [18] with the three different DVFs: fixed-size BMA, unweighted variable-size BMA and perceptually weighted variable-size BMA. Results regarding two standard sequences, *Taxi* and *Rubik*, are considered here because they represent extreme cases for the segmentation algorithm: small moving objects with simple translations and one large object rotating around an axis, respectively. While in the first sequence even simple velocity clustering could solve the problem, in the second one, regions with very different velocities have to be merged in a single complex-motion object.

Table II shows the average segmentation error (normalized by the maximum in each sequence), the number of iterations and the number of identified regions in each case. Figs. 10 and 11 show examples of the motion flows and the segmentations achieved in particular frames of the sequences along with the corresponding segmentation error. The differences in robustness, coherence and meaningfulness between the results of the different algorithms are apparent in the motion flows (Figs. 10 and 11). The quantitative results of the segmentation confirm this intuitive impression. The block appearance of the layers is obviously due to the sparseness of the input flows. The segmentation algorithm is usually initialized with dense flows (one motion vector per pixel) [17], [18], not with sparse (one motion vector per block) flows. However, this worst case situation is appropriate to highlight the usefulness of each flow.

The following trends can be identified from the obtained results: 1) the segmentation is better when the blocks are fairly small compared to the size of the moving regions (see the variable-size BMA error results for *Rubik* –large object- and *Taxi* –small objects–); however, 2) the segmentation is very sensitive to the robustness and coherence of the sparse flow. This implies that the block size cannot be arbitrarily reduced to improve the segmentation resolution: Despite its higher density, a more noisy flow from a too small block size BMA is worse.

This general behavior may be explained in terms of the data needed to estimate the six parameters of the affine models. Ideally, a minimum of three independent vectors (six data) are needed to segment a region. In real (noisy) situations, more vectors per region will be needed. Consequently, the segmentations of large objects are comparatively better. On the other hand, as the number of motion measurements is reduced when using sparse flows (compared to dense flows), their robustness becomes critical. This is why the results based on fixed-size BMA (sparse and noisy) are extremely poor. However, if the variable-size BMA used is robust enough, reliable rough segmentations are still possible. In particular, the increased robustness of the perceptually weighted flow speeds up the convergence of the segmentation algorithm and minimizes the

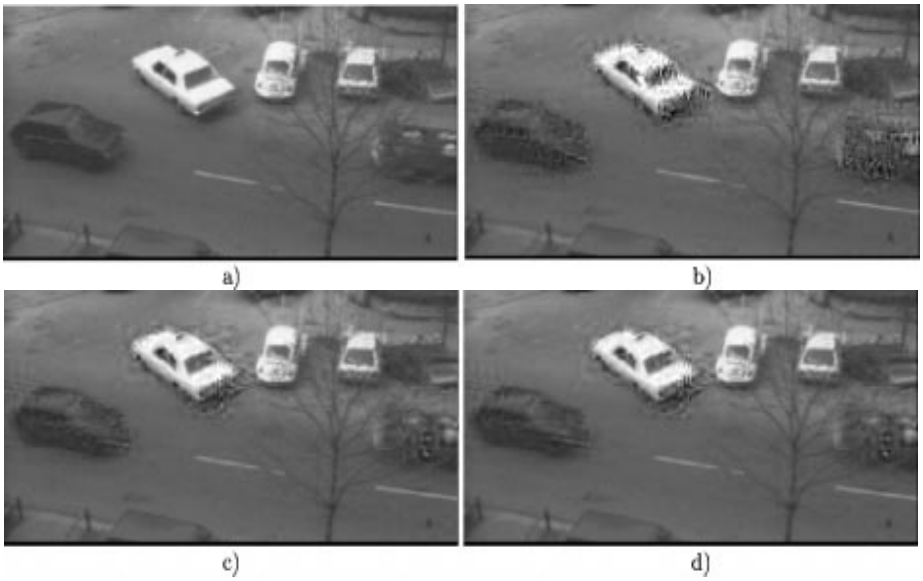


Fig. 8. Reconstruction results with different motion estimations and a fixed MPEG-like quantization. (a) Original (detail of frame 7 of the *Taxi* sequence). (b) Fixed-size BMA. (c) Unweighted variable-size BMA. (d) Weighted variable-size BMA.

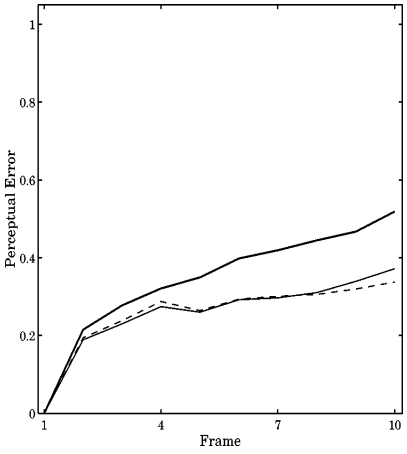


Fig. 9. Perceptual distortion measures for the frames of the *Taxi* sequence using different motion estimations with the same (MPEG-like) quantizer. The thick solid line corresponds to the fixed-size BMA. The thin solid line corresponds to the unweighted variable-size BMA and the thin dashed line corresponds to the weighted variable-size BMA.

segmentation error (see Table II). Moreover, a closer estimation of the actual number of moving objects is obtained due to the smoothness of the flow within moving regions, which prevents false splittings.

To summarize, despite the fact that a better bit allocation between DVF and DFD (Table I and Fig. 7) does not improve the quality of the decoded sequence (Figs. 8 and 9), the segmentation results (Table II and Figs. 10 and 11) show that perceptual weighting gives rise to a more meaningful flow: it makes the sparse flow more suitable for obtaining robust rough segmentations.

C. Experiment 3: Relative Relevance of the Improvements in the Motion Estimation and the Quantization

To study the combined effect and the relative advantages of the proposed improvements, the four possible combinations of motion estimation and quantization algorithms were compared at a fixed bit-rate. The perceptually weighted and the unweighted (suboptimal) variable-size BMA were combined with the the 2-D linear (MPEG-like) and the 2-D nonlinear MPE quantizers. Fig. 12 shows an example of a reconstructed frame using the four different approaches considered. Fig. 13 shows the corresponding distortion for each frame of the *Rubik* sequence. These reconstruction examples and error results confirm what expected from the separate analysis of the previous sections: The quality improvement due to the use of the appropriate entropy measure in the motion estimation is negligible compared to the benefits of a better quantization.

In this case, the simple increase in the quantizer bandpass that comes from considering the visual nonlinearities (Fig. 2) may explain the enhancement. Due to the noisy (high-frequency) nature of the error signal, wide band quantizers may be better than narrower band (CSF-based) quantizers. In fact, the benefits of this enhanced quantizer are more apparent on video frames (Figs. 5 and 12) than in still images [13], [14]. The interesting constant distortion result reported in Section V-A (Fig. 6) is also reproduced here: While the perceptual distortion increases quickly when using the linear quantizer, it remains constant with the nonlinear MPE quantizer regardless the motion estimation algorithm.

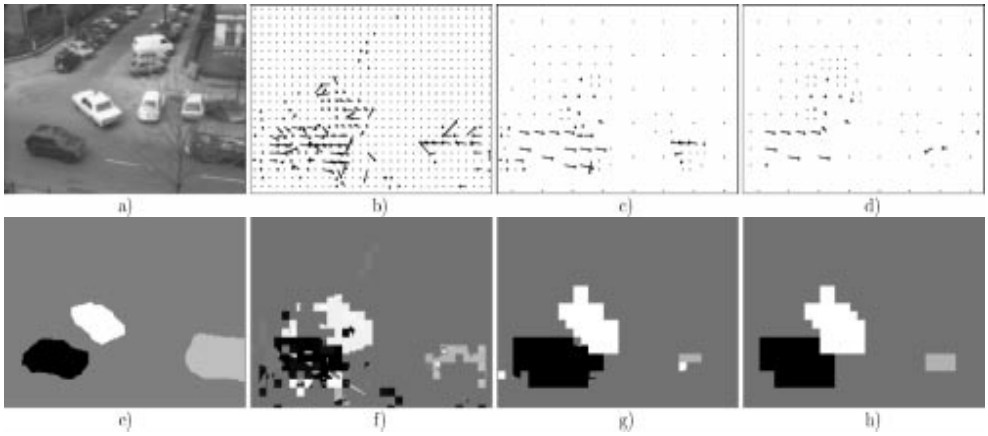


Fig. 10. DVFs and motion-based segmentation results (*Taxi*, frame 9). (a) Original frame. (b) Fixed-size BMA flow. (c) Unweighted variable-size BMA flow. (d) Perceptually weighted variable-size BMA flow. (e) Ideal segmentation. (f) Segmentation with fixed-size BMA ($\epsilon = 1.02$). (g) Segmentation with unweighted variable-size BMA, ($\epsilon = 0.49$). (h) Segmentation with perceptually weighted variable-size BMA, ($\epsilon = 0.35$).

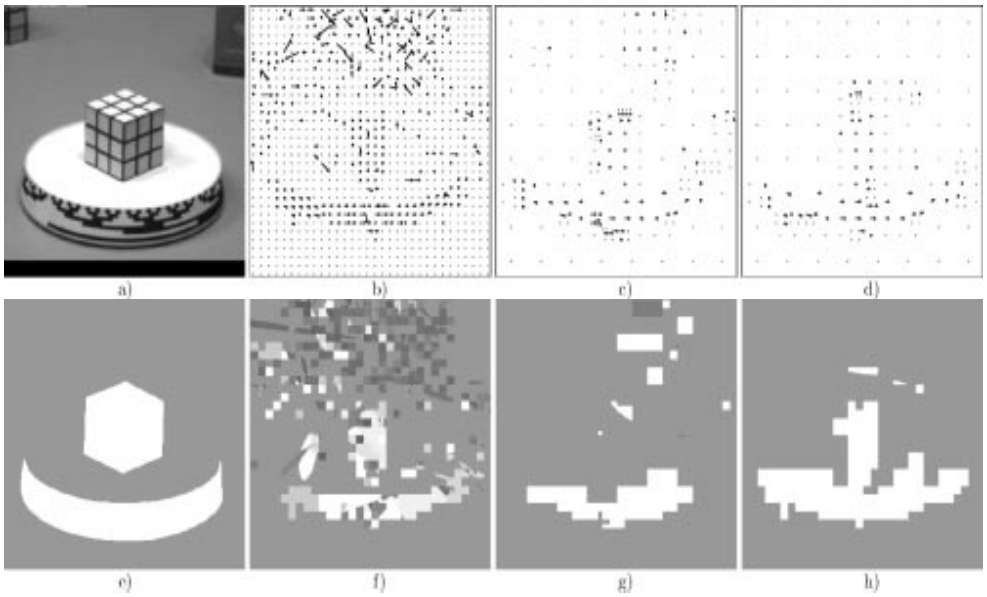


Fig. 11. DVFs and motion-based segmentation results (*Rubik*, frame 3). (a) Original frame. (b) Fixed-size BMA flow. (c) Unweighted variable-size BMA flow. (d) Perceptually weighted variable-size BMA flow. (e) Ideal segmentation. (f) Segmentation with fixed-size BMA ($\epsilon = 0.98$). (g) Segmentation with unweighted variable-size BMA, ($\epsilon = 0.51$). (h) Segmentation with perceptually weighted variable-size BMA, ($\epsilon = 0.25$).

TABLE II
SEGMENTATION ERRORS, NUMBER OF ITERATIONS AND IDENTIFIED REGIONS

	RUBIK (2 objects)			TAXI (4 objects)		
	ϵ	Iterat.	Regions	ϵ	Iterat.	Regions
FSBMA	1.00 ± 0.05	25	21.0 ± 4.8	1.00 ± 0.14	25	15.8 ± 3.8
Unweight. VSBMA	0.46 ± 0.06	21 ± 5	4.4 ± 2.2	0.52 ± 0.15	11 ± 9	5.8 ± 2.0
Weighted VSBMA	0.27 ± 0.03	7 ± 6	2.1 ± 0.3	0.49 ± 0.14	4 ± 1	4.3 ± 0.5

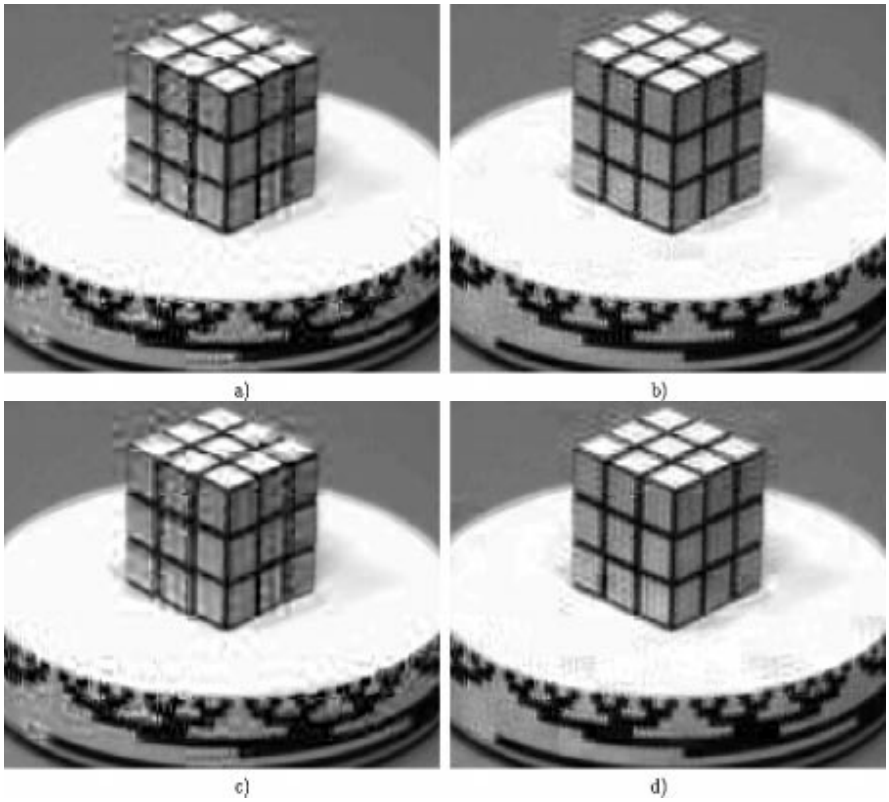


Fig. 12. Decoded results using different combinations of quantizers and motion estimations. (a) Unweighted variable-size BMA and 2-D linear MPE, MPEG-like, quantization. (b) Unweighted variable-size BMA and 2-D nonlinear MPE quantization. (c) Perceptually weighted variable-size BMA and 2-D linear MPE, MPEG-like, quantization. (d) Perceptually weighted variable-size BMA and 2-D nonlinear MPE quantization.

D. Experiment 4: The Proposed Scheme Versus Previous Comparable Schemes

In this section, the considered elements of the motion compensated video coder were combined to simulate and compare previously reported schemes (MPEG-1 or H.261 and H.263) and the proposed ones. MPEG-1 and H.261 use a fixed-size BMA and a linear (CSF-based) MPE quantizer. A regular implementation of H.263 use an unweighted splitting criterion variable-size BMA and a linear MPE quantizer. The video coder scheme proposed here use a perceptually weighted variable-size BMA and a nonlinear MPE quantizer (either 2-D or 3-D).

Examples of the decoded frames at a fixed bit-rate are shown in Figs. 14 and 15. The perceptual distortion for each frame of the reconstructed sequences is shown in Fig. 16. Both frames and distortions confirm the improvements of the proposed schemes compared to the previous similar schemes in terms of subjective quality at a fixed bit-rate.

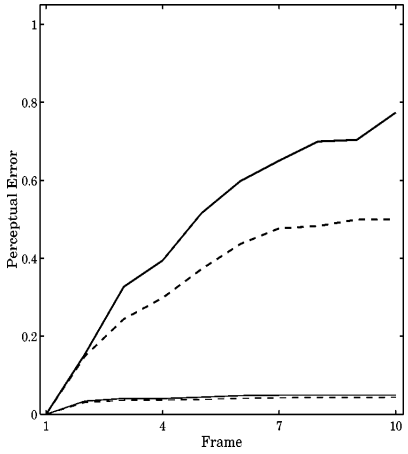


Fig. 13. Perceptual distortion measures for the frames of the *Rubik* sequence using different combinations of quantizers and motion estimations. The thick lines correspond to the approaches using the linear MPE quantizer. The thin lines correspond to the approaches using the 2-D nonlinear MPE quantizer. The dashed lines correspond to the approaches using perceptually weighted variable-size BMA and the solid lines correspond to the approaches using unweighted variable-size BMA.

VI. Final Remarks

The current motion compensated video coding standards include very basic perceptual information (linear threshold models) only in the quantizer design. In this paper a multigrid motion compensated video coding scheme based on a more

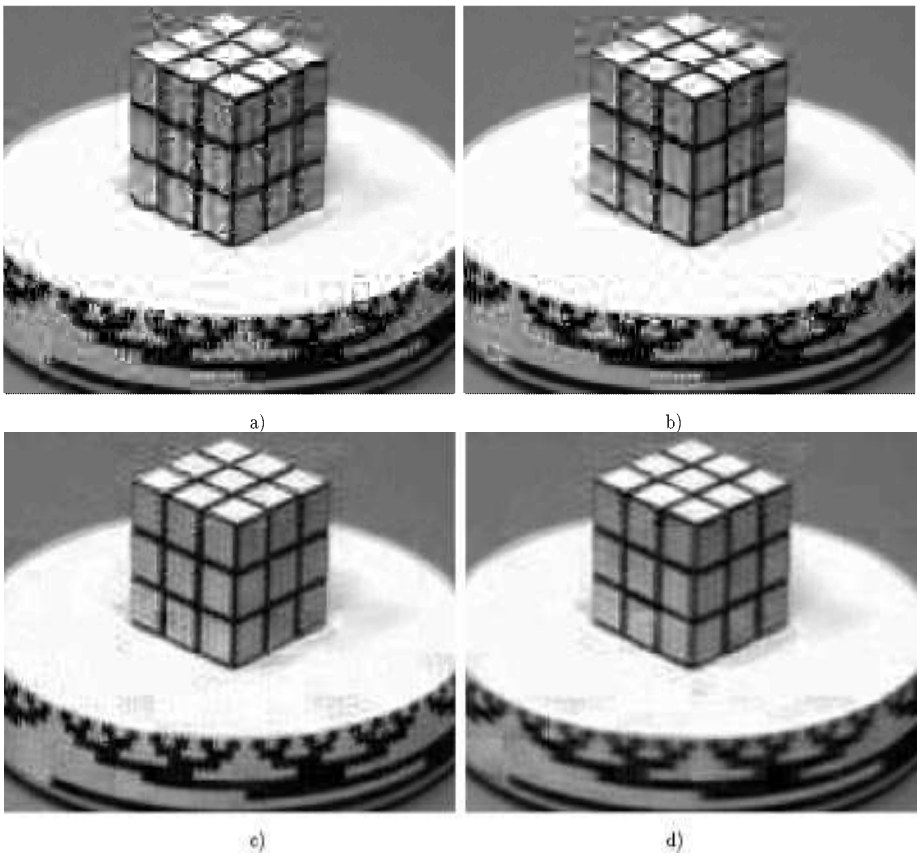


Fig. 14. Reconstruction results with the *Rubik* sequence using previously reported encoding configurations (a-b) and the proposed 2-D or 3-D alternatives (c-d). (a) Fixed size BMA for motion estimation and MPEG-like quantization (linear MPE). (b) Unweighted variable-size BMA and MPEG-like quantization (linear MPE). (c) Perceptually weighted variable-size BMA and 2-D nonlinear MPE quantization. (d) Perceptually weighted variable-size BMA and 2-D nonlinear MPE quantization and temporal filtering.

accurate HVS contrast discrimination model has been presented. The model accounts for the nonuniform nature of the HVS redundancy removal in the frequency domain. Here the basic idea is to design the entire encoding process to preserve no more than the subjectively significant information at a given subjective distortion level. This aim affects not only the quantizer design, but also the motion estimation.

On the one hand, as a result of a more accurate perception model and the MPE restriction criterion, an improved nonlinear quantizer has been proposed. On the other hand, this perceptual quantizer is used here to decide if additional motion information is perceptually significant. This definition of perceptually significant motion information gives rise to an appropriate entropy-constrained BMA (using the actual DFD entropy) in a natural way. In this way, superfluous effort in the motion description (predicting details that are going to be discarded by the quantizer) is avoided and a perceptual feedback is introduced in the motion estimation refinement.

The reconstructed frames and perceptually meaningful distortion measures show that the proposed schemes improve the results of previous comparable schemes such as H.263 with unweighted motion refinement and MPEG-like quantization. In particular, nonlinear MPE quantizers lead to better subjective quality than the linear MPE (CSF-based, MPEG-like) quantizers at the same bit-rates because they more accurately preserve the relevant information of the DFDs. A very interesting consequence is that the (2-D or 3-D) nonlinear quantizers keep the distortion bounded over a large group of frames, so the need to introduce bit-consuming intracoded frames can be reduced. According to the perceptual distortion results at a fixed bit-rate, the benefits of a better bit allocation between DVF and DFD (the benefits of the perceptual weight in motion estimation) are negligible compared to the benefits of perceptually more accurate quantizers. The significant improvement due to this enhancement of the quantizer suggests that quantizer design may be more important than optimal motion estimation from the rate-distortion point of view.

However, a side effect of the perceptual control of the motion estimation is a scale-dependent refinement strategy that gives rise to more robust and meaningful motion flows compared to unweighted refinement criteria. This is confirmed by the segmentation results that show how the perceptual weighting

Fig. 15. Reconstruction results with the *Taxi* sequence using previously reported encoding configurations (a-b) and the proposed 2-D or 3-D alternatives (c-d). (a) Fixed size BMA for motion estimation and MPEG-like quantization (linear MPE). (b) Unweighted variable-size BMA and MPEG-like quantization (linear MPE). (c) Perceptually weighted variable-size BMA and 2-D nonlinear MPE quantization. (d) Perceptually weighted variable-size BMA and 2-D nonlinear MPE quantization and temporal filtering.

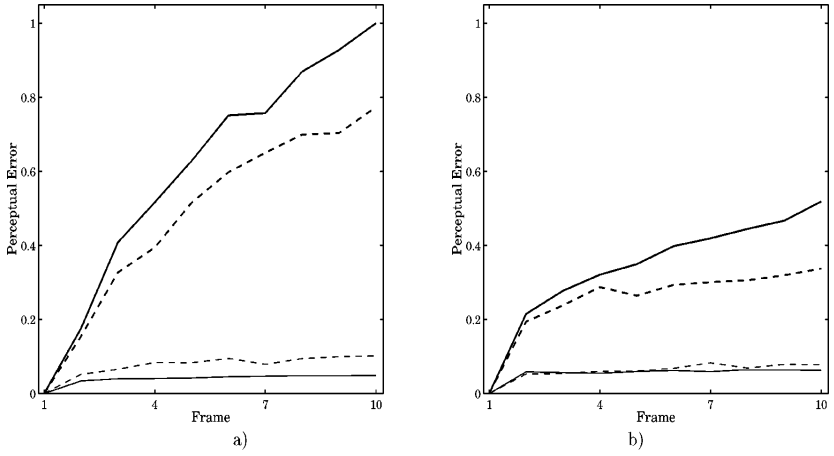


Fig. 16. Perceptual distortion measures for the frames of the (a) *Rubik* and (b) *Taxi* sequence using previous and the proposed schemes. The thick lines correspond to the previous apporaches and thin lines correspond to the proposed schemes. Solid thick line corresponds to fixed-size BMA and linear MPE quantizer (H.261, MPEG1). Dashed thick line corresponds to unweighted variable-size BMA and linear MPE quantizer (H.263). The solid thin line corresponds to perceptually weighted variable-size BMA and 2-D nonlinear MPE and the dashed thin line corresponds to perceptually weighted variable-size BMA and 3-D nonlinear MPE. As in Fig. 6, the frame-by-frame implementation of the perceptual distortion measure may slightly overestimate the visual effect of the frame blurring introduced by the temporal filter.

makes a sparse flow suitable to obtain rough segmentations of the moving objects. These results suggest that the perceptual quantizer should not to be taken into account in the motion estimation due to rate-distortion reasons, but to obtain more meaningful flows that may be of interest for higher-level video coding.

Systematic psychophysical testing instead of the simple frame-by-frame perceptual measure may be necessary to measure accurately the relevance of the temporal features of the proposed 1-D+2-D quantizer design. Nevertheless, the results presented suggest that when amplitude nonlinearities are taken into account, the consideration of the temporal properties is not so significant.

VII. Acknowledgements

The authors wish to thank the reviewers for their detailed comments on the original manuscript. The authors also wish to thank Dr. J. Soret for valuable discussions on the hardware implementation of the algorithms described in this paper.

References

[1] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, no. 4, pp. 47–58, 1991.

[2] G. Tziritas and C. Labit, *Motion Analysis for Image Sequence Coding*, Amsterdam, The Netherlands: Elsevier, 1994.

[3] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[4] *ITU-Telecommunication standardization sector*, Draft recommendation H.263, 1994.

[5] *Overview of the MPEG-4 Version 1 Standard*, ISO/IEC JTCI/SC29/WG11 N1909, 1997.

[6] M. Bierling, "Displacement estimation by hierarchical block-matching," *Proc. SPIE, Conf. Vis. Commun. Image Process.*, vol. 1001, pp. 942–951, 1988.

[7] F. Moscheni, F. Dufaux, and H. Nicolas, "Entropy criterion for optimal bit allocation between motion and prediction error information," *Proc. SPIE, Conf. Vis. Commun. Image Process.*, vol. 2094, pp. 235–242, 1993.

[8] F. Dufaux and F. Moscheni, "Motion estimation techniques for digital TV: A review and new contribution," *Proc. IEEE*, vol. 83, no. 6, pp. 858–876, 1995.

[9] G. M. Schuster and A. K. Katsaggelos, *Rate-Distortion Based Video Compression*. Norwell, MA: Kluwer, 1997.

[10] ——, "A video compression scheme with optimal bit allocation among segmentation, motion and residual error," *IEEE Trans. Image Processing*, vol. 6, no. 11, pp. 1487–1502, 1997.

[11] J. Lee, "Joint optimization of block size and quantization for quadtree-based motion estimation," *IEEE Trans. Image Processing*, vol. 7, pp. 909–912, June 1998.

[12] J. Malo, F. Ferri, J. Albert, and J. M. Artigas, "Splitting criterion for hierarchical motion estimation based on perceptual coding," *Electron. Lett.*, vol. 34, no. 6, pp. 541–543, 1998.

[13] J. Malo, F. Ferri, J. Albert, and J. Soret, "Comparison of perceptually uniform quantization with average error minimization in image transform coding," *Electron. Lett.*, vol. 35, no. 13, pp. 1067–1068, 1999.

[14] J. Malo, F. Ferri, J. Albert, J. Soret, and J. M. Artigas, "The role of perceptual contrast nonlinearities in image transform coding," *Image Vis. Comput.*, vol. 18, no. 3, pp. 233–246, 2000.

[15] J. Malo, A. M. Pons, and J. M. Artigas, "Subjective image fidelity metric based on bit allocation of the human visual system in the DCT domain," *Image Vis. Comput.*, vol. 15, no. 7, pp. 535–548, 1997.

[16] A. M. Pons, J. Malo, J. M. Artigas, and P. Capilla, "Image quality metric based on multidimensional contrast perception models," *Displays*, vol. 20, pp. 93–110, 1999.

[17] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Processing*, vol. 3, pp. 625–638, May 1994.

[18] L. Torres, D. García, and A. Mates, "A robust motion estimation and segmentation approach to represent moving images with layers," in *IEEE ICASSP*, Munich, Germany, 1997, pp. 2981–2984.

[19] G. E. Legge, "A power law for contrast discrimination," *Vis. Res.*, vol. 18, pp. 68–91, 1981.

[20] D. H. Kelly, "Motion and vision II: Stabilized spatiotemporal threshold surface," *J. Opt. Soc. Amer.*, vol. 69, no. 10, pp. 1340–1349, 1979.

[21] B. L. Beard, S. Klein, and T. Carney, "Motion thresholds can be predicted from contrast discrimination," *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 14, no. 9, pp. 2449–2470, 1997.

[22] E. Martinez-Uriegas, "Color detection and color contrast discrimination thresholds," in *Proc. OSA Annu. Meeting ILS-XIII*, Los Angeles, CA, 1997, p. 81.

[23] B. Girod, "Motion compensation: Visual aspects, accuracy and fundamental limits," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds., 1993.

[24] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. IEEE*, vol. 81, no. 10, pp. 1385–1422, 1993.

[25] A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Hum. Vis., Vis. Proc. Digit. Disp. IV*, vol. 1913, 1993.

[26] J. Malo, A. M. Pons, and J. M. Artigas, "Bit allocation algorithm for codebook design in vector quantization fully based on human visual system nonlinearities for suprathreshold contrasts," *Electron. Lett.*, vol. 31, no. 15, pp. 1222–1224, 1995.

[27] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.

[28] B. Macq, "Weighted optimum bit allocations to orthogonal transforms for picture coding," *IEEE J. Select. Areas Commun.*, vol. 10, no. 5, pp. 875–883, 1992.

[29] M. Chan, Y. B. Yu, and A. G. Constantinides, "Variable size bock matching motion compensation with applications to video coding," *Proc. IEEE, Vis., Image, Sig. Process.*, vol. 137, no. 4, pp. 205–212, 1990.

[30] E. Reusens *et al.*, "Dynamic approach to visual data compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 197–211, 1997.

[31] F. Dufaux and F. Moscheni, "Segmentation-based motion estimation for second generation video coding techniques," in *Video Coding: A Second Generation Approach*, L. Torres and M. Kunt, Eds. Norwell, MA: Kluwer, 1996.

[32] F. Dufaux, I. Moccagatta, B. Rouchouze, T. Ebrahimi, and M. Kunt, "Motion-compensated generic coding of video based on multiresolution data structure," *Opt. Eng.*, vol. 32, no. 7, pp. 1559–1570, 1993.

[33] J. Li and X. Lin, "Sequantial image coding based on multiresolution tree architecture," *Electron. Lett.*, vol. 29, no. 17, pp. 1545–1547, 1993.

[34] A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 14, pp. 2379–2391, 1997.

[35] H. R. Wilson, "Pattern discrimination, visual filters and spatial sampling irregularities," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, Eds. Cambridge, MA: MIT Press, 1991, pp. 153–168.

[36] A. B. Watson, "Efficiency of a model human image code," *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 4, no. 12, pp. 2401–2417, 1987.

[37] J. G. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," *IEEE Trans. Biomed. Eng.*, vol. 36, pp. 107–114, 1989.

[38] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *Proc. SPIE*, vol. 1666, pp. 365–374, 1992.

[39] J. Solomon, A. Watson, and A. Ahumada, "Visibility of DCT basis functions: effects of contrast masking," in *Proc. Data Compress. Conf.*, Snowbird, UT, 1994, pp. 361–370.

[40] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 2, pp. 127–135, 1982.

[41] K. N. Nygan, H. C. Koh, and W. C. Wong, "Hybrid image coding scheme incorporating human visual system characteristics," *Opt. Eng.*, vol. 30, no. 7, pp. 940–946, 1991.

[42] D. J. Heeger, "Model for the extraction of image flow," *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 4, pp. 1455–1471, 1987.

[43] A. B. Watson and A. J. Ahumada, "Model of human visual motion sensing," *J. Opt. Soc. Amer. A Opt. Image Sci.*, vol. 2, pp. 322–342, 1985.

[44] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.

[45] W. Niehsen and M. Brünig, "Covariance analysis of motion-compensated frame differences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 536–539, 1999.

[46] S. Winkler, "Issues on vision modeling for video quality assessment," *Signal Process.*, vol. 78, no. 2, pp. 231–252, 1999.

[47] A. B. Watson, J. Hu, J. F. McGowan, and J. B. Mulligan, "Design and performance of a digital video quality metric," *Proc. SPIE, Hum. Vis., Vis. Proc. Digit. Disp. IX*, vol. 3644, pp. 168–174, 1999.
[48] G. Borgefors, "Distance transformations in digital images," *Comput. Vis. Graph. Image Process.*, vol. 34, pp. 344–371, 1986.

Jesús Malo was born in 1970. He received the M.Sc. degree in physics in 1995 and the Ph.D. degree in physics in 1999, both from Universitat de València, València, Spain.

Since 1994, he has been with the Vision Group of Universitat de València. In the Fall of 1999, he worked with the Image and Vision Group at the Institute of Optics, CSIC. In 2000 and 2001, he worked as Fulbright Postdoctoral Fellow at the Vision Group of the NASA Ames Research Center, Moffett Field, CA, and at the Laboratory of Computational Vision of the Center for Neural Science, New York University. He is interested in models of low-level human vision, their relations with information theory and their applications to computer vision, image processing, and vision science experimentation.

Dr. Malo was the recipient of the Vistakon European Research Award in 1994.

Juan Gutiérrez was born in 1971. He received the Licenciado degree in physics (electricity, electronics, and computer science) in 1995 from the Universtitat de València, València, Spain, where he is currently pursuing the Ph.D. degree in motion estimation and segmentation.

Since 1997, he has been with the Computer Science and Electronics Department, Universtitat de València, València, Spain, as a Teacher and a Research Fellow. He has performed two research stays, one at the Digital Imaging Research Centre at Kingston University (U.K.) for seven months working on multi-object tracking, and the other at the Department of Informatics and Mathematical Modeling, Technical University of Denmark for two months, working on optical flow regularization. His current research interests include image analysis, motion understanding, and regularization theory.

I. Epifanio, photograph and biography not available at time of publication.

Francesc J. Ferri was born in 1964. He received the Licenciado degree in physics (electricity, electronics, and computer science) in 1987 and the Ph.D. degree in pattern recognition in 1993, both from Universitat de València, València, Spain.

He has been with the Computer Science and Electronics Department, Universtitat de València, since 1986; first as a Research Fellow, and since 1988, as a Teacher of computer science and pattern recognition. He has been involved in a number of scientific and technical projects on computer vision and pattern recognition. During a sabbatical in 1993, he joined the Vision, Speech, and Signal Processing Research group at the University of Surrey (U.K.), where he was working in feature selection and statistical pattern recognition methodology. He has authored or coauthored more than 90 technical papers in international conferences and well-established journals in his fields of interest. He has also helped in refereeing several of these publications. His current research interests include feature selection, nonparametric classification methods, inductive learning, computational geometry, image analysis, and motion understanding.

Dr. Ferri is a member of the ACM and the AERFAI (Spanish Association for Pattern Recognition and Image Analysis, affilate society of IAPR), where he has helped several times in organizing the biennal Spanish Conference on Pattern Recognition and Image Analysis.

José M. Artigas received the M.S. degree in physical sciences in 1974 and the Ph.D. degree in 1979, both from Universitat de València, València, Spain.

From 1985 to 1992, he worked in collaboration with the group of Prof. F. W. Campbell of Cambridge University (U.K.). Since 1984, he has been on the staff of the Faculty of Physics, Universitat de València, where he has held the position of Associate Profesor of Optics. His two main research interests have been in physiological optics (spatial vision) and color vision. He has advised nine Ph.D. dissertations. He is currently developing a project concern a new noninvasive methods for early detection of visual pathologies.

Dr. Artigas is a member of the Spanish Optical Society (SEDO) and the American Optical Society (OSA).

E. TRABAJO 5

Regularization operators for natural images based on non-linear perception models.

Juan Gutiérrez, Francesc J. Ferri y Jesús Malo

IEEE Transactions on Image Processing (aceptado en noviembre de 2004)

Regularization Operators for Natural Images based on Non-Linear Perception Models

Juan Gutiérrez(*), Francesc J. Ferri and Jesús Malo

***Abstract*— Image restoration requires some a priori knowledge of the solution. Some of the conventional regularization techniques are based on the estimation of the power spectrum density. Simple statistical models for spectral estimation just take into account second order relations between the pixels of the image. However, natural images exhibit additional features, such as particular relationships between local Fourier or wavelet transform coefficients. Biological visual systems have evolved to capture these relations.**

We propose the use of this biological behavior to build regularization operators as an alternative to simple statistical models. The results suggest that if the penalty operator takes these additional features in natural images into account, it will be more robust and the choice of the regularization parameter is less critical.

***Index Terms*— Regularization, image restoration, early vision models, natural image statistics.**

I. Introduction

IMAGES are produced to record or display useful information regarding some phenomenon of interest. When the process of image formation and recording is not perfect, only a degraded version, $i_d(x)$, of the original image, $i_o(x)$, is available. The whole distortion process is usually modelled by means of a linear operator acting upon the original image, plus additive noise:

$$i_d(x) = h(x, x') \otimes i_o(x') + n(x) \tag{1}$$

The purpose of image restoration is to estimate the original image from the distorted and noisy one. This is a non-trivial problem because even if the operator is known, a naive inversion will amplify the noise [1].

Multiple techniques of different nature have been proposed to solve this problem. Most of them require (and critically depend on) some image model or some *a priori* assumptions about the image or the noise [2], [3], [1]. The most exploited feature of natural images for building an image model is the specific low-pass behavior of the power spectrum [4], [5]. This feature can be easily described in the spatial domain by an autocorrelation function or the coefficients of an autoregressive (AR) model [1], [6]. In more recent approaches, the image model includes additional features. In particular, certain relationships between neighboring coefficients in a transformed domain are considered [7], [8].

In the context of regularization techniques, the restored image is a trade-off between closeness to the observed data and the absence of some *unwanted features* in the solution [9]. The problem is again to use an appropriate image model to design the regularization operator so as to preserve the *suitable features* in the solution.

In this work we put forward and analyze a class of penalty operators [10] for image regularization which are not based on a model of the image or the noise, but on the current model of low-level human visual perception [11], [12]. In this way, there is no need to make any *a priori* statistical assumptions.

The use of perception models to design regularization operators relies on the Barlow hypothesis [13], in which he argued that biological vision systems have evolved for an optimal processing of natural images. A number of results support this hypothesis. First, it has been shown that the early processing mechanisms in the visual cortex perform a linear wavelet-like transform using a set of filters similar to those obtained by applying *Independent Component Analysis* (ICA) to a set of natural images [14]. Second, biological vision systems exhibit non-linear interactions between the responses of the linear wavelet-like stage. In these non-linearities each coefficient is normalized by a combination of neighboring coefficients [11], [12], [15]. It has been shown that this behavior gives rise to signals with increased independence (see [15], [16], or [17]. By maximizing independence, biological visual systems capture the most salient features of natural images and, in a way, discriminate between useful (*suitable*) and negligible (*unwanted*) information.

From this ecological point of view, current models of early vision implicitly include the generic statistical properties of natural images (as for instance their rough spectral properties and the relationships between the coefficients in a local frequency domain [15], [17]). Therefore, it makes sense to use the perceptual response in the regularization context for restoration purposes, because it can tell us what features should be present in the solution and what features should be removed from it.

The use of simple human vision properties in regularization is not new [1], [2]. However, these approaches used simpler (linear) models and, of greater importance, the rationale was very application-oriented (similar to the one which is popular in coding literature [18], [19]): as the images are addressed to human viewers, features which are not subjectively relevant should be penalized.

JG (juan.gutierrez@uv.es) and FJF are with the Department d'Informàtica, Universitat de València. JM (jesus.malo@uv.es) is with the Department d'Optica, Universitat de València, Doctor Moliner, 50. 46100 Burjassot (Valencia). Spain. JG and JM are members of the VISTA lab at the Universitat de València (http://www.uv.es/vista/vistavalencia/).

This work has been supported by the grants TIC2003-01504 (CICYT), UV-AE-20030394 (Universitat de València), Grupos04-08 (Generalitat Valenciana), and by the program DN8644.

Apart from using of a more advanced model, the main idea of this work is different: we want to stress the fact that the information about natural images implicitly included in the human visual response can be used as a successful alternative to explicit image models in the penalty operator design. Moreover, if the penalty operator takes into account these additional features in natural images, it may be more robust and the choice of the regularization parameter is less critical. It is important to stress that the statistical benefits of using this advanced perceptual model are not limited to its application in restoration: see [17] for an application in image coding.

This paper is organized as follows: in Section II we review the regularization framework. In Section III we stress that the $\frac{1}{f}$ spectrum shape is not the only common feature shared by natural images. In particular, we show that the coefficients of local Fourier representations are related. These relationships are equivalent to those recently reported in other local frequency domains [7], [8], [15], [17], [20], [21], [22]. In Section IV we propose the use of the current non-linear early vision response to define regularization operators. We show that this response implicitly takes into account both band limitation and the additional features illustrated in Section III. Our experimental results appear in Section V, followed by our concluding remarks.

II. Regularization and image features

In the Tikhonov regularization framework [9] the image estimate, i', is the one that simultaneously minimizes both the departure from the observed data, $f_1(h(x,x') \otimes i(x'), i_d(x))$, as well as a penalty functional, $f_2(p(x,x') \otimes i(x))$, which measures the presence of some *unwanted feature* in the solution:

$$i'(x) = \arg\min_{i(x)} f_1(h(x,x') \otimes i(x'), i_d(x)) + \lambda^2 f_2(p(x,x') \otimes i(x')) \quad (2)$$

where $p(x,x')$ is a linear operator that extracts the features that should be minimized in the solution, and λ, referred to as the regularization parameter, establishes the trade-off between both terms. If Euclidean distance is used to measure these departures, f_1 and f_2, we have [9],

$$i'(x) = \arg\min_{i(x)} \big(|h(x,x') \otimes i(x') - i_d(x)|^2 + \lambda^2 |p(x,x') \otimes i(x')|^2 \big) \quad (3)$$

Many different choices exist for f_1, f_2 and the penalty operator $p(x,x')$. Common choices for the measures are l_p-norms [23], usually l_1 and l_2-norms. In particular, the effect of a small exponent ($p < 2$) in f_2 relatively decreases the penalization of large variations in luminance, thus allowing the presence of sharp discontinuities in the solution [24]. On the other hand, the penalty operator is usually taken as a fixed high-pass filter (e.g. first or second derivative) or can be adaptively computed [1]. In this work we propose an adaptive operator and analyze its behavior in the particular case of Eq. 3.

Assuming a local approach in Eq. 3, h and p may be considered shift invariant, and hence $\otimes$ is a convolution, therefore the solution in a region x_0 of a block-Fourier domain is [2]:

$$I'(x_0,f) = \frac{H^*(x_0,f)}{|H(x_0,f)|^2 + \frac{\lambda(x_0)^2}{|P(x_0,f)^{-1}|^2}} \, I_d(x_0,f) \quad (4)$$

The meaning of the regularization operator in the Fourier domain, $P(x,f)$, in terms of a simple image model, can be inferred by analogy from the result using the optimal Wiener filter [2]:

$$I'(x_0,f) = \frac{H^*(x_0,f)}{|H(x_0,f)|^2 + \frac{|N(x_0,f)|^2}{|I_o(x_0,f)|^2}} \, I_d(x_0,f) \quad (5)$$

In the Wiener solution the inverse operator $H^*/|H|^2$ is modified using information about the power spectrum of the original signal and the noise (which are not available in general). Eq. 5 suggests that the penalty operator, P, and the parameter, λ, should be related to the power spectrum density of the signal and the energy of the noise respectively. Therefore, the conventional operator design is usually based on a model of the autocorrelation of the image [6], [1]. The smoothness assumption (limited band signal) gives rise to high-pass penalty operators, so, in this case the *unwanted feature* is the energy in the high frequency region.

A critical issue in the regularization framework is the choice of the regularization parameter, λ. This problem gets worse if the operator is not specifically matched to the signal. For instance, when a generic high-pass operator is used, the appropriate restoration of high frequency components in the signal becomes very *sensitive* to λ. However, the importance of the regularization parameter may be reduced if the operator not only includes second-order relations but also some other features which are common in natural images.

III. Additional features in natural images beyond $\frac{1}{f}$ spectrum

Natural images present a strong correlation between the luminance of pixels in the spatial domain. This is the reason for assuming Gaussian models for natural images [25], [26], [20]. Under the Gaussian assumption, a simple second order approach fully describes the signal. This simple characterization has been successfully exploited in applications where removing the statistical dependence of the samples is required [27], [28]. Principal Component Analysis (PCA) on a set of natural images gives rise to Fourier-like eigenvectors, i.e., oscillating functions extending all over the spatial domain. Moreover, the energy (the square of the eigenvalues) is concentrated in the low-frequency PCA components [25], [26], [20]. Specifically, natural images exhibit a $\frac{1}{f^\gamma}$ power spectrum [4]. This justifies the use of high-pass regularization operators (such as first and second derivatives of the signal [1]) in image restoration.

However, natural images are not that simple. In order to consider higher order interactions, ICA techniques [29], [30] have been developed as an alternative to PCA. When applied to natural images, wavelet-like representations emerge, i.e., spatially localized oscillating functions. Nevertheless, when representing images in these local frequency domains, not all higher order interactions are taken into account [7], [15], [17], [20], [21], [22]. Specifically, intriguing relationships among the coefficients of these linear frequency transforms still remain. Linear transforms cannot remove all the statistical dependence because natural images do not come from a linear combination of signals from independent sources (e.g. due to occlusion). Modelling these statistical relations led to improvements in image coding applications [21], [31].

Instead of modelling these relations explicitly, in this work we are going to take them into account by using a non-linear perceptual model [11], [12]. The use of these perception models rather than the explicit statistical modelling of the relations in a local Fourier domain has been successfully applied in image coding applications (see [32], [33], or [17]).

In this section we describe a way to show these relations in the particular local frequency domain which will be used to define the regularization operators. This is intended to clarify how the perceptual response reviewed in next section takes these relationships into account.

We have computed the cross-correlation between the absolute value of the 32×32 block-Fourier coefficients of the widely-used natural images database [34]. As the local Fourier spectrum of natural images is not stationary, a direct comparison between coefficients at very different frequencies is biased. As stated above, natural images exhibit a $\frac{1}{f^\gamma}$ power spectrum, meaning that, the comparison of a high frequency coefficient with a low frequency coefficient is biased by the high energy of the latter. Therefore, instead of a direct comparison, we first divide each coefficient by $\frac{1}{f}$ (assuming that $\gamma \approx 2$ [5]). In that way, a unit mean process is obtained and a fair computation of the cross-correlation can be done.

Figure 1 shows an estimate of the cross-correlation for two particular coefficients. Obviously, for any given coefficient the cross-correlation is maximum for itself and its symmetrical. When compared to other coefficients, the cross-correlation decreases as the distance in frequency increases. The relationships illustrated in this example are consistent with those reported in other local frequency transform domains (see [7], [8], [15], [20], [21], [22], or [17]).

In summary, natural images have a low-pass spectrum (as assumed by conventional statistical models used for regularization), but they also exhibit the particular behavior illustrated by these relationships in a local frequency domain. This additional feature of natural images should not be neglected.

IV. Perceptual Regularization Operators

The standard model of human low-level image analysis [11], [12], [15] has two basic stages in which the input image, $i(x)$, is first transformed into a vector, $I(x,f)$, in a local frequency domain (using a linear filter bank, T) and then it is non-linearly transformed into a response vector, $r(x,f)$,

$$i \xrightarrow{T} I \xrightarrow{R} r$$

A local (block) Fourier transform will be used here as a model of the perceptual linear transform T. The response to each coefficient of the transform, $R(I(x_0,f))$, is given by a non-linearity in which the energy of each coefficient is normalized by a linear combination of the neighbors (*masking* behavior). This non-linear response is commonly referred to as *divisive normalization* [12], [11]. For a particular region, x_0, of a local Fourier transform, the response for each frequency component is:

$$r(x_0,f) = R(I(x_0,f)) = \frac{\alpha(f)|I(x_0,f)|^2}{\beta(f) + (k(f,f') \otimes |I(x_0,f')|^2)} \quad (6)$$

where the parameters $\alpha(f)$, $\beta(f)$ and $k(f,f')$, are chosen to fit the psychophysical response of human viewers to periodic functions [16]. The vector, r, is constituted by the responses of the set of perceptual mechanisms that analyze the image. Eq. 6 describes the fact that the response of a particular mechanism tuned to (x_0,f), not only depends on the energy of the linear output, $I(x_0,f)$, but is also influenced by its neighbors, $I(x_0,f')$, with $f' \neq f$ [11], [12], [15]. In particular, the sensitivity of the mechanism tuned to (x_0,f) is decreased when the input image excites high responses in the neighboring coefficients of the linear stage. The kernel, $k(f,f')$, determines the extent of this neighborhood in frequency and orientation. It is usually assumed to be Gaussian so that the influence of closer neighbors is bigger [11], [12], [15]. Figure 2 shows the frequency dependence of the parameters in Eq. 6 and some examples of the response for three particular mechanisms. The effect of the denominator in Eq. 6 is to saturate the response for high energy inputs in the local frequency region determined by k: such input stimuli *mask* the response of the considered mechanism. Note that the overall behavior of the response is mediated by the band-pass filter, $\alpha(f)$, which is similar to the classical *Contrast Sensitivity Function* (CSF) [35].

As stated in the introduction, the responses in Eq. 6 capture the most salient independent features of natural images and discriminate between useful and negligible information from the perceptual point of view. Therefore, the perceptual

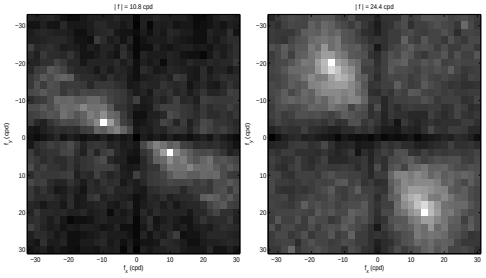


Fig. 1. Empirical interaction of two particular coefficients of the local Fourier Transform with their neighbors. The absolute value of the frequency of these two coefficients is $|f| = 10.8$ and $|f| = 24.4$ cycles/degree (cpd). A sampling frequency of 64 cpd is assumed.

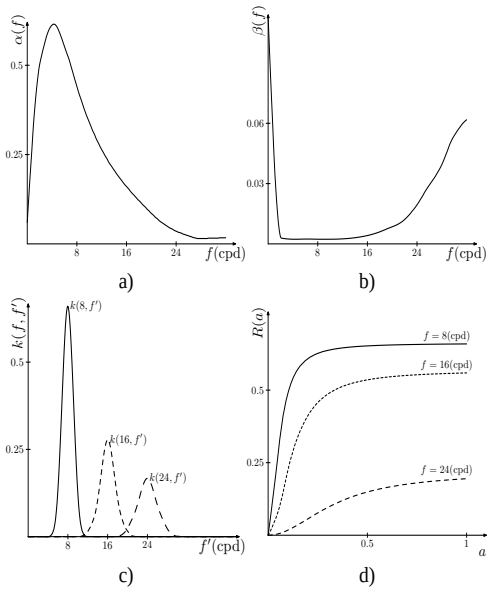


Fig. 2. Parameters of the vision model and non-linear response functions. The values in figure 2.d assume that the Fourier coefficients (denoted by a) are expressed in contrast (amplitude over mean luminance). The responses of figure 2.d show the basic sigmoid behavior of Eq.6, but they are just particular examples for isolated sinusoids. In general these response curves will depend on the neighbor coefficients. Note that the parameters are slightly different from those reported in [17] because we are using here a local Fourier transform instead of a local DCT and a slightly different model. However, the final behavior (figure 2.d) is the same.

responses for a particular region of the observed data may tell us what features should be present in the solution and what features can be removed from it. This can be seen as a perceptual spectral estimation adapted to the most common features present in natural images. According to this perception-based heuristic, we may use the response Eq. 6 to estimate the (perceptually relevant) spectrum in Eq. 5. We therefore propose to define the regularization operator, P, of Eq. 4 as inversely proportional to the response. Thus, each restored region in the local frequency domain, $I'(x_0, f)$, will be obtained as,

$$I'(x_0, f) = \frac{H^*(x_0, f)}{|H(x_0, f)|^2 + \frac{\lambda^2}{|R(I_d(x_0, f))|^2}} \, I_d(x_0, f) \qquad (7)$$

Previous simpler perceptual approaches based on the linear CSF model [2] can be seen as a particular case of the more general class of operators presented here.

The perception model just described includes the behavior exhibited by natural images, as illustrated in Section III. Note that the Gaussian kernels in Figure 2.c follow the trend shown in Figure 1: the interaction neighborhood increases with frequency. Therefore, the response implicitly includes two facts: rough low-pass behavior through the function $\alpha(f)$, and the particular relationships between the energy of neighboring coefficients through the kernels $k(f, f')$.

From the point of view of the Barlow hypothesis, the goal of the divisive normalization is to make those parts of the signal that cannot be predicted from the neighborhood more explicit. In addition, it has been reported that this energy normalization acts as a sort of divisive DPCM [16]: the energy in each transform coefficient, $I(f)$, is predicted from the energy of the coefficients in the Gaussian neighborhood, $k(f, f')$. This is how the saliency of the relevant coefficients of the linear transform is increased in the response [15], [16].

This predictive behavior is what makes normalization interesting for regularization operator design. Using the analogy between Eqs. 5 and 7, we see that the role of the response model is the extraction of the relevant features of the image to estimate the spectrum of the original signal from its degraded version. Relevant features are those that cannot be predicted from the neighborhood. Wide-band noise introduces two non-natural features: a noticeable change in the shape of the spectrum and increased predictability in the frequency domain. This is true if the bandwidth of the noise is bigger than the kernels in Eq. 6. In this case the energy of the noise can be predicted from the energy of the neighbors. This represents a potential limitation of the proposed operators because they would consider a high energy noise of too narrow a bandwidth as a relevant feature.

As in this framework an adaptive penalty operator has to be obtained from degraded images, the operator should be robust to noise. *Robustness* in this work is understood as the ability to preserve the same components of the signal that would be preserved with an ideal operator obtained from the original signal. The predictive behavior of the divisive normalization may be useful to achieve robustness. If an image is degraded by adding wide-band noise, the contribution of the noise to each transform coefficient can be predicted from the neighboring coefficients. In this way, the proposed operator will detect the relevant components despite the noise.

A real example of the ability of the response, r, to capture relevant features in the presence of noise is shown in Figure 3. Figure 3.a shows a block taken from the image Barbara with a texture (a salient feature at a particular frequency). The same block blurred and corrupted with white noise is shown in Figure 3.b. Figures 3.c and 3.d show the corresponding regularization operators using the proposed method together with the standard second-derivative operator. As discussed above, the relevant feature at frequency $f \approx 18$ cpd is preserved despite the noise, while a second-derivative operator penalizes this feature.

V. Results and Discussion

In this section the properties of the proposed operator will be compared to other penalty operators taking three aspects into account: a) the sensitivity of the solution to λ, b) the robustness of the operators and, c) their performance in the restoration task using a variety of synthetic degradations. Finally, we also analyze the performance of the proposed method with naturally degraded images.

Five standard images have been used for the experiments: Barbara, Lena, Einstein, boats and peppers. We assume that these 256×256 images have a physical extent of 4×4 degrees, i.e. the sampling frequency is 64 cpd. The proposed

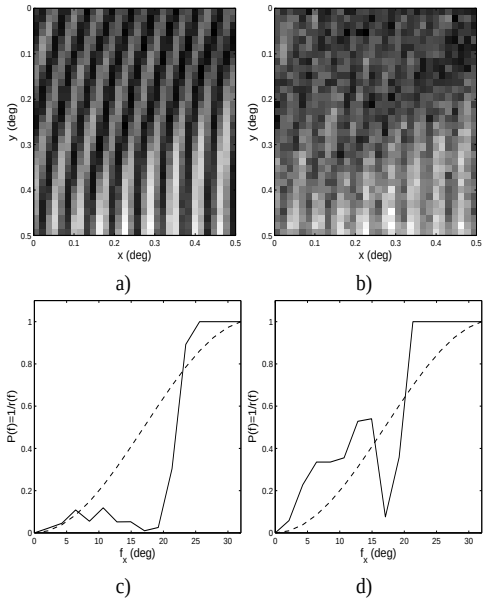


Fig. 3. Perceptual penalty operators (solid curves) in (c,d) obtained from a particular block without noise (a) and with noise (b), compared with the operator based on the second derivative (dashed curves). Notice how, with the proposed approach, the relevant feature (texture of 18 cpd) is not penalized.

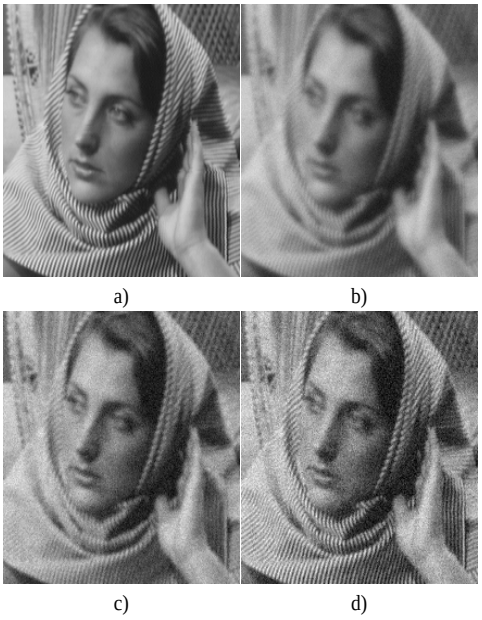


Fig. 4. Zoom of original and representative degraded images. a) Original Barbara image, b) Distorted image (almost deblurring problem): $f_c = 3.2$ cpd, $\sigma_n^2 = 15$, c) Distorted image: $f_c = 16$ cpd, $\sigma_n^2 = 100$, and d) Distorted image (almost denoising problem): $f_c = 27.2$ cpd, $\sigma_n^2 = 200$.

regularization operator has been tested in all images with different degradations. The images were blurred using known linear shift invariant low-pass filters with normalized cutoff frequencies of 0.1, 0.5 and 0.85 (3.2, 16 and 27.2 cpd) and were corrupted with additive white Gaussian noise of different variances (from $\sigma_n^2 = 5$ to $\sigma_n^2 = 400$). These restoration problems range from almost pure deblurring (very low cutoff frequency and low energy noise) to almost pure denoising (high cutoff frequency and high energy noise). Figure 4 shows illustrative examples of these degradations using the standard image Barbara.

The different restoration methods have been applied to 50 realizations of the above mentioned corruption processes over each original image. In the experiments, the regularization parameter, λ, has been obtained using the standard L-curve method [36] for each block of the local Fourier Transform. In all cases, the borders of neighboring 16×16 blocks have been overlapped to reduce the blocking artifacts common to all methods.

Euclidean distortion measures may not take into account subjective aspects related to the empirical Mean Opinion Score given by an average human observer [37]. Therefore, a perceptually meaningful metric (Square Root Integral (SQRI) measure [38]) was also used along with the standard Peak Signal-to-Noise Ratio (PSNR). However, the same trends revealed by PSNR were obtained SQRI, as can be seen by subjective inspection of the images (which is the really important issue).

In this work we will compare the proposed operator with other operators based on spectral estimation, so that they can be used in Eq. 4. In particular, we considered three AR models [1], [6] of increasing neighborhoods as shown in Figure 5. The coefficients of these AR models have been locally fitted to capture the non-stationarity of the signal. From those coefficients a local estimation of the power spectrum is obtained [39], to be used in Eq. 4.

Additionally, two classical spatially invariant operators are included for reference purposes: the second derivative and the CSF-based operator reported in [40]. On the one hand, the second derivative operator captures the generic $\frac{1}{f}$ spectral behavior of natural images. In fact, when AR models are trained on data sets which are large enough, operators which are quite similar to the second derivative are obtained. On the other hand, the CSF-based operator represents a simple example of the use of perceptual information in regularization.

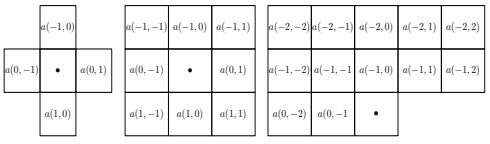


Fig. 5. Neighborhoods of the particular 2D autoregressive models used here: AR(4) (left), AR(8) (center), and AR(12) (right). The dotted pixel is linearly predicted from the corresponding neighborhood weighted by $a(m, n)$.

A. Sensitivity to the regularization parameter

As discussed in Section II, the sensitivity of the solution to the regularization parameter depends on the ability of the operator to capture the properties of the signal. Roughly speaking, a solution is more sensitive if small variations in the regularization parameter lead to big changes in the quality of the solution.

Given a particular operator, P, a solution, $\hat{i}_P(\lambda)$, can be computed for each value of λ. A particular λ selection method, such as the L-curve [36], obtains a particular value λ_L, which leads to a particular solution $\hat{i}_P(\lambda_L)$. Assuming that the original image, i_o, is available, and given a distortion measure $d(\cdot,\cdot)$, the optimal value of λ could be computed as,

$$\lambda_{opt} = \arg\min_{\lambda} D_P(\lambda)$$

where $D_P(\lambda) = d(i_o, \hat{i}_P(\lambda))$. The sensitivity of the solution to λ when a particular operator P is used, can be defined as the change in the image distortion with regard to the departure from the optimal λ_{opt}. That is,

$$S(P) = \frac{\Delta D_P}{\Delta\lambda} = \frac{|D_P(\lambda_{opt}) - D_P(\lambda_L)|}{|\lambda_{opt} - \lambda_L|}$$

The meaning of this sensitivity measure is illustrated in Figure 6.

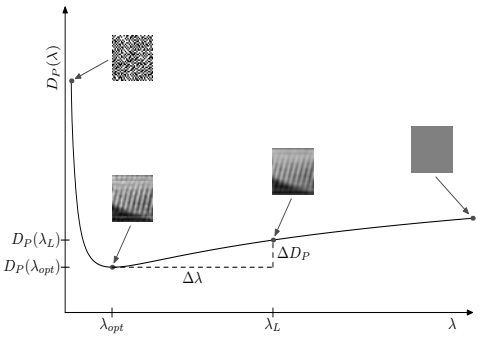


Fig. 6. Distortion of the set of solutions $\hat{i}_P(\lambda)$ for a particular operator P. The solution ranges from noise ($\lambda <<$) to over-smoothed images ($\lambda >>$). Representative points have been highlighted. Note that this sensitivity is related to the concavity of this curve which depends on P.

The particular sensitivity values considering different operators have been measured at each block of the Barbara image using both Mean Square Error (MSE) and SQRI. A standard boxplot (omitting outliers for clarity) of the corresponding results for the MSE case is shown in Figure 7. Similar results are obtained when SQRI is used. The proposed operator and the CSF lead to smaller and less scattered sensitivity values. As a consequence, using perceptual operators makes the choice of λ less critical.

B. Robustness of the operators

As discussed in Section IV, robustness in this work is understood as the ability of an adaptive operator to preserve the same components of the signal that would be preserved with an ideal operator obtained from the original signal. Thus, this property only applies to adaptive operators.

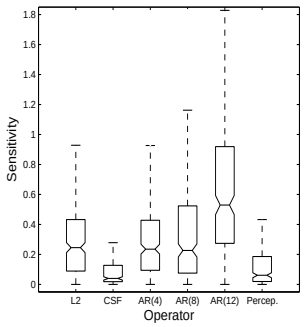


Fig. 7. Sensitivity of the different operators using image Barbara.

If the original image, i_o, is available, any adaptive operator P estimated from the degraded image can be compared to the same operator P_o obtained from the original one. The corresponding MSE between the operators can be used to assess their similarity. Consequently, we define the robustness of a particular adaptive operator as,

$$R(P) = \text{MSE}(P, P_o)$$

The robustness of the different adaptive operators considered for the Barbara image is shown as a boxplot in Figure 8. As the proposed operator captures the complexity of the signal better, it is more robust to the presence of degradations.

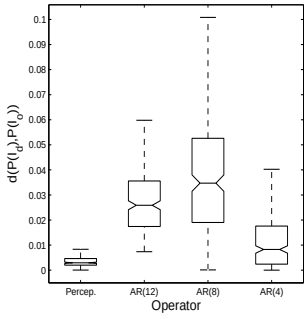


Fig. 8. Robustness of the different adaptive operators.

C. Restoration results

Two different sets of restoration tasks have been considered to assess the performance of the proposed method: first different synthetic degradations, and second naturally degraded images.

Figures 10-12 show representative restored images corresponding to the set of operators considered in this work applied to the synthetic degraded images. These images correspond to the restoration of the degraded Barbara images shown in

Figure 4. Tables I-III show the distortion results for all the experiments carried out. These results are the average over the five considered images. Figure 9 shows the details of one of the distorted images used to compute the averages in Tables I-III. As can be seen, the variances in the results due to the 50 realizations of the degradation are small. Similar plots are obtained for the remaining images and degradations.

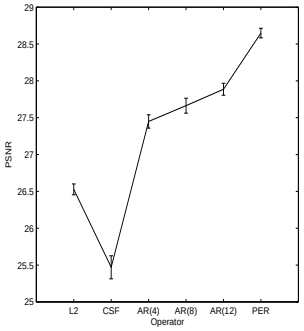


Fig. 9. Example of the results of a particular restoration experiment. Average distortions and standard deviations over 50 realizations of the degradation for a particular image (Barbara) and a particular combination of cutoff frequency (16 cpd) and variance of the noise ($\sigma_n^2 = 75$)

From these results, it is obvious that the performance of spatially invariant approaches (sub-figures *a* and *b* in Figures 10-12) is worse than the performance of the adaptive ones (sub-figures *c-f* in Figures 10-12). In particular, spatially invariant approaches give rise to over-smoothed areas in all cases. Similar results are obtained when other stationary models for the autocorrelation are used (e.g. autoregressive models trained with a wide image data set).

As expected, adaptive approaches are superior to stationary ones. If the complexity of the model is increased, AR(4) to AR(8), the power spectrum is more accurately estimated. Consequently, a better adaptation to each block is obtained and the performance improves (Table I-III). This trend is broken if the complexity of the AR model keeps increasing, because over-fitting occurs. This effect is particularly noticeable when the energy of the noise is high (Table III and Figure 12, where the AR(12) model is learning the noise).

On the other hand, the proposed approach constitutes a trade-off between the overall noise and the preservation of specific details, as can be observed in sub-figure *f* in Figures 10-12. The artifacts obtained with this kind of penalty operators are different from those obtained using the autoregressive methods. Our approach removes wide band noise because it is predictable (in the local Fourier domain) from the neighbors. This means that this wide band noise is not considered to be a relevant feature. However, unpredictable features in the transform domain, as for instance salient textures, are preserved as discussed in Section IV.

The differences in performance are less noticeable in the first experiment (almost a deblurring problem, see Table I and Figure 10). In this case, the low energy of the noise makes the problem easier and all the adaptive methods perform

TABLE I

PSNR FOR DIFFERENT OPERATORS AND NOISE LEVELS FOR 5 IMAGES FOR A BLURRING WITH A CUTOFF FREQUENCY OF 3.2 CPD.

	$\sigma_n^2 = 5$	$\sigma_n^2 = 10$	$\sigma_n^2 = 15$
L2	28.55	28.74	28.55
CSF	29.25	28.49	28.04
AR(4)	30.26	29.62	29.08
AR(8)	30.66	29.76	29.23
AR(12)	30.44	29.50	29.02
PER	**30.71**	**29.75**	**29.20**

TABLE II

PSNR FOR DIFFERENT OPERATORS AND NOISE LEVELS FOR 5 IMAGES FOR A BLURRING WITH A CUTOFF FREQUENCY OF 16 CPD.

	$\sigma_n^2 = 25$	$\sigma_n^2 = 50$	$\sigma_n^2 = 75$	$\sigma_n^2 = 100$
L2	30.60	29.49	28.82	28.33
CSF	30.04	29.33	28.76	28.25
AR(4)	30.53	29.66	29.04	28.51
AR(8)	30.78	29.78	29.19	28.67
AR(12)	31.09	29.98	29.23	28.63
PER	**31.15**	**30.43**	**29.87**	**29.34**

similarly. However, when the energy of the noise is increased, the differences are more noticeable, as can be seen in the last experiment (almost a denoising problem, see Table III and Figure 12). The proposed operator works better than the other operators considered in a wider range of degradations, which is consistent with the sensitivity and robustness results described above.

Finally, the proposed method is applied to the restoration of naturally degraded images. The pictures were taken using a SONY DSC P8 digital camera. In this experiment, all the images were on focus, so that the degradation comes from eventual poor illumination conditions (unknown photon noise), and from the artifacts due to the (unknown) processing in the acquisition and storage of the images (eg. demosaicing, compression, etc.). Therefore, in this case, the proposed method is compared to a standard denoising method [41] [1].

In realistic conditions the original image is not available in general, so the computation of quantitative distortions is not straightforward. In order to give indicative distortion measures, we took two versions of each image: the first one was taken in good illumination conditions, and the second one was taken reducing the numerical aperture or the exposure time so that the integrated irradiance at the CCD was reduced. The histogram of the second image was linearly corrected

[1]We used the Matlab implementation of Lee's algorithm in the function `wiener2.m`

TABLE III

PSNR FOR DIFFERENT OPERATORS AND NOISE LEVELS FOR 5 IMAGES FOR A BLURRING WITH A CUTOFF FREQUENCY OF 27.2 CPD.

	$\sigma_n^2 = 100$	$\sigma_n^2 = 200$	$\sigma_n^2 = 300$	$\sigma_n^2 = 400$
L2	29.37	28.00	27.05	26.19
CSF	29.14	28.40	27.77	27.20
AR(4)	29.62	28.14	27.00	26.20
AR(8)	30.26	28.92	27.98	27.04
AR(12)	27.16	23.98	22.59	21.40
PER	**30.67**	**29.77**	**29.23**	**28.364**

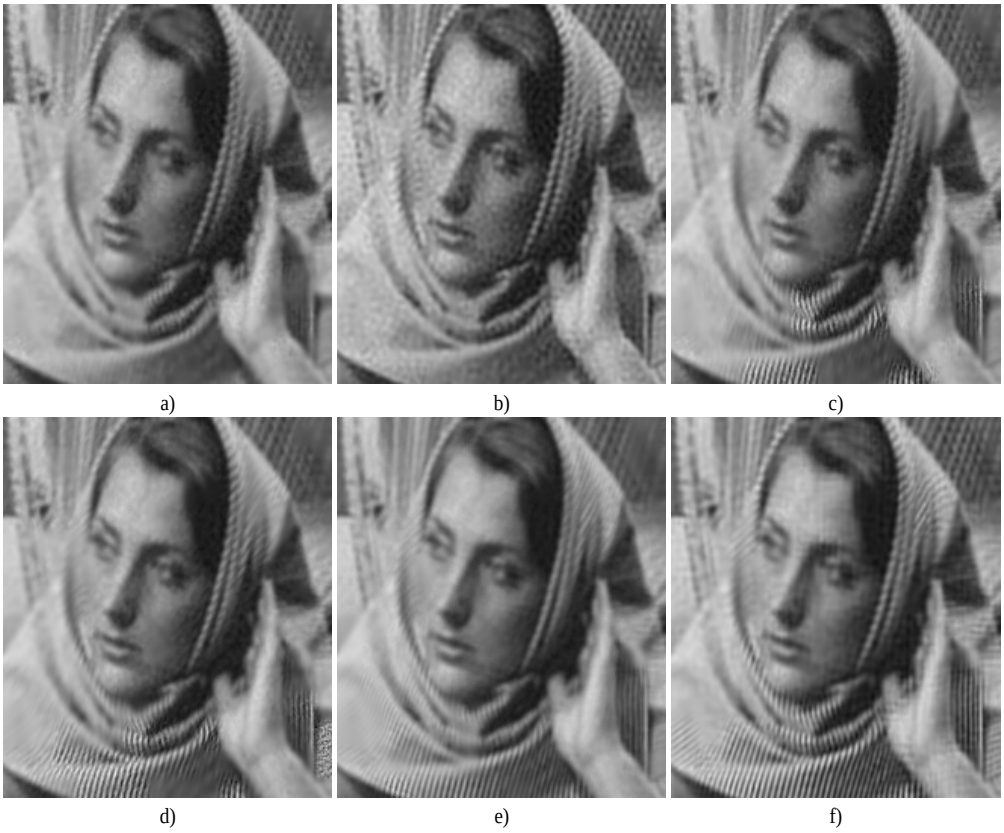


Fig. 10. Restoration results for the degraded image in figure 4.*b* ($f_c = 3.2$ and $\sigma_n^2 = 15$). a) Second-derivative based result, b) CSF-based result, c) Image restored using adaptive AR(4), d) Image restored using adaptive AR(8), e) Image restored using adaptive AR(12), f) Image restored using the proposed method.

to reproduce the mean and variance (average luminance and contrast) of the first one. In this way the second image is a naturally distorted version of the first image. We will refer to the first one as *reference image* and to the second one as *naturally degraded image*.

Figures 13-15 show representative results of the above distortion procedure in a range of illuminations and the corresponding results obtained using Wiener denosing [41] and the proposed method. Subplots e and f in these figures show the difference between the restored and the degraded images. This difference represents the part of the degraded image identified as noise by each method. It is worth noting that while the standard algorithm removes features that may be relevant (letters, edges and structures of the objects), the proposed algorithm preserves these structures focusing in removing the noise.

The above qualitative results are consistent with the distortion measures given in table IV. As can be seen, both restoration techniques improve the estimated PSNR, and the proposed method provides a better result using this measure. However, it has to be stressed that these measures are just an approximation to the true PSNR because they have been computed using a reference image which is not the (unknown) original.

TABLE IV
ESTIMATED PSNR FOR THE NATURALLY DEGRADED IMAGES.

	Device	Stones	Building
Degraded	20.80	28.14	23.39
Wiener	25.33	28.45	23.96
PER	**25.64**	**29.07**	**24.08**

VI. CONCLUSIONS AND FURTHER WORK

In this paper, the information about natural images implicitly included in the current model of early human vision (approximate $\frac{1}{f}$ behavior and relations between coefficients in a local frequency transform domain) has been used to define regularization operators for restoration purposes. This class of operators shows several advantages with regard to the adaptive operators considered based on AR models: they are less sensitive to the regularization parameter, more robust to the degradation and give rise to less noticeable artifacts than the other methods. Besides, in naturally degraded images they

Fig. 11. Restoration results for the degraded image in figure 4.*c* ($f_c = 16$ and $\sigma_n^2 = 100$). a) Second-derivative based result, b) CSF-based result, c) Image restored using adaptive AR(4), d) Image restored using adaptive AR(8), e) Image restored using adaptive AR(12), f) Image restored using the proposed method.

preserve the relevant structure of objects when compared with a standard Wiener denoising.

These results suggest that the proposed method, which does not require statistical *a priori* assumptions on the image, can be used as a successful alternative to operators based on explicit models of the spectrum of the image.

The proposed operator has been used in a l_2 norm penalty functional. An interesting extension of this work could explore the behavior of the proposed operator with more general norms. For instance, using the l_1 norm with the proposed operator would improve edge preservation, as already reported using stationary operators such as the first derivative [24].

ACKNOWLEDGMENTS

We thank Prof. P.C. Hansen for fruitful comments on the L-curve method [36]. We thank Prof. E. Simoncelli for his criticism on the used transform: even though the block Fourier transform is nice enough to illustrate the reported behavior, the better way to do it is using his Steerable Pyramid [15]. We thank Dr. G. Camps for their useful discussions on autoregressive models. We thank Dr. Viru Garcia for his insight on natural images features. We thank Prof. J.M. Artigas, Dr. A.B. Watson and Dr. A.M. Pons for their comments on the psychophysical meaning of the kernels $k(f, f')$ [12]. We thank Dr. P. Mulet for leting us to use his Total Variation code [42] during the reviewing process.

REFERENCES

[1] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Sig. Proc. Mag.*, vol. 14, pp. 24–41, 1997.
[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. NJ: Prentice Hall, 1977.
[3] A. K. Katsaggelos, "Iterative image restoration algorithms," *Opt. Eng.*, vol. 28, no. 7, pp. 735–748, 1989.
[4] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Physical Review Letters*, vol. 73, no. 6, pp. 814–817, 1994.
[5] D. C. Knill, D. Field, and D. Kersten, "Human discrimination of fractal images," *Journal of the Opt. Soc. of Am.*, vol. 7, no. 6, pp. 1113–1123, 1990.
[6] R. Molina, J. Núñez, F. J. Cortijo, and J. Mateos, "Image restoration in astronomy," *IEEE Sig. Proc. Mag.*, pp. 11–29, March 2001.
[7] E. P. Simoncelli, "Statistical models for images: Compression, restoration and synthesis," in *31st Asilomar Conf. on Signals, Systems and Computers*. Pacific Grove, CA: IEEE Computer Society, 1997, pp. 673–678.
[8] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Trans. on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

Fig. 12. Restoration results for the degraded image in figure 4.*d* ($f_c = 27.2$ and $\sigma_n^2 = 200$). a) Second-derivative based result, b) CSF-based result, c) Image restored using adaptive AR(4), d) Image restored using adaptive AR(8), e) Image restored using adaptive AR(12), f) Image restored using the proposed method.

[9] M. Bertero, T. A. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, no. 8, pp. 869–889, 1988.

[10] J. Gutierrez, J. Malo, and F. Ferri, "Perceptual regularization functionals for natural image restoration," in *Proc. IEEE ICIP*, vol. 2, 2003, pp. 989–992.

[11] M. Carandini and D. Heeger, "Summation and division by neurons in visual cortex," *Science*, vol. 264, no. 5163, pp. 1333–6, 1994.

[12] A. Watson and J. Solomon, "A model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America A*, vol. 14, pp. 2379–2391, 1997.

[13] H. B. Barlow, "What is the computational goal of the neocortex?" in *Large scale neuronal theories of the brain*, C. Koch, Ed. MIT Press, 1994, pp. 1–22.

[14] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[15] O. Schwartz and E. P. Simoncelli, "Natural signal statistics and sensory gain control," *Nature Neuroscience*, vol. 4, no. 8, pp. 819–825, 2001.

[16] J. Malo, R. Navarro, I. Epifanio, F. Ferri, and J. Artigas, "Non-linear invertible representation for joint statistical and perceptual feature decorrelation," *Lect. Not. Comp. Sci.*, vol. 1876, pp. 658–667, 2000.

[17] J. Malo, I. Epifanio, R. Navarro, and E. Simoncelli, "Non-linear image representation for efficient perceptual coding," *To appear in: IEEE Transactions on Image Processing*, 2005.

[18] G. Wallace, "The JPEG still picture compression standard," *Comm. ACM*, vol. 34, no. 4, pp. 31–43, 1991.

[19] J. Malo, J. Gutierrez, I. Epifanio, F. Ferri, and J. Artigas, "Perceptual feed-back in multigrid motion estimation using an improved dct quantization," *IEEE Trans. Im. Proc.*, vol. 10, no. 10, pp. 1411–1427s, 2001.

[20] E. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, 2001.

[21] R. Buccigrossi and E. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Proc.*, vol. 8, no. 12, pp. 1688–1701, 1999.

[22] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Transactions on Image Processing*, vol. 10, no. 11, pp. 1647–1658, 2001.

[23] Y. Li, "A globally convergent method for l_p problems," *SIAM J. Optimization*, vol. 3, no. 3, pp. 609–629, 1993.

[24] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[25] R. Clarke, "Relation between the Karhunen-Loeve transform and cosine transforms," *Proceedings IEE, F*, vol. 128, no. 6, pp. 359–360, 1981.

[26] P. J. B. Hancock, R. J. Baddeley, and L. S. Smith, "The principal components of natural images," *Network: Computation in Neural Systems*, vol. 3, no. 1, pp. 61–70, 1992.

[27] A. Gersho and R. Gray, *Vector quantization and signal compression*. Boston: Kluwer Academic Publishers, 1992.

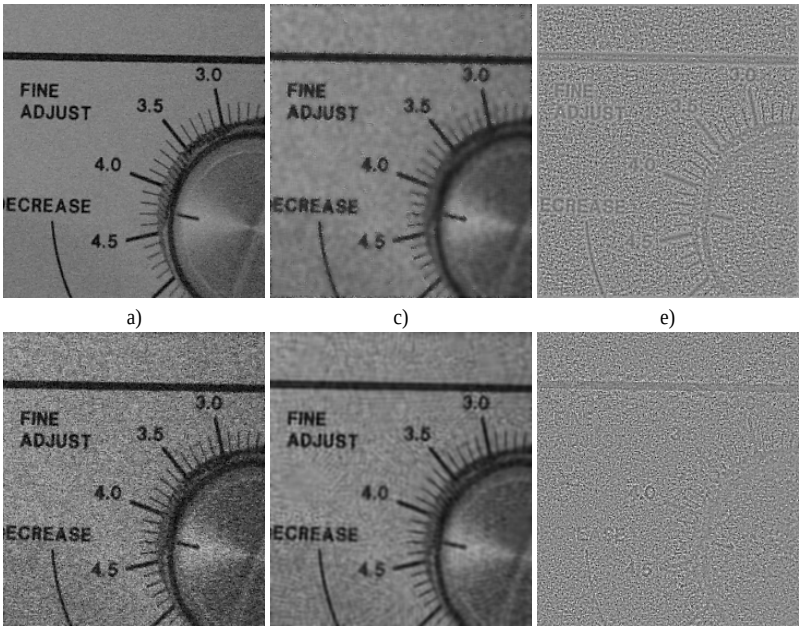
[28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2000.

[29] J. Cardoso, "Source separation using higher order moment," in *Int. Conf. Acoustics Speech Signal Proc.*, 1989, pp. 2109–12.

[30] P. Comon, "Independent component analysis: A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[31] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy image compression," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1303–1310, 1996.

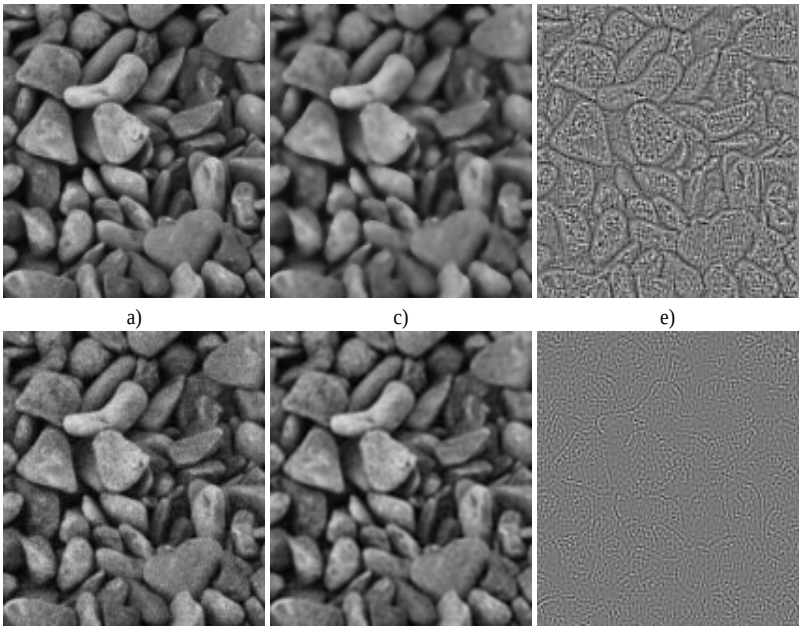
[32] I. Epifanio, J. Gutierrez, and J. Malo, "Linear transform for simultaneous



a) c) e)

b) d) f)

Fig. 13. Restoration results for a naturally degraded image (Device). a) Reference image (exposure time 1/6 sec, aperture N=5.2), b) Naturally degraded image (exposure time 1/25 sec, aperture N=5.2) , c) Restored image using Wiener denoising, d) Restored image using the proposed method, e) Information removed by the Wiener method, f) Information removed by the proposed method.



a) c) e)

b) d) f)

Fig. 14. Restoration results for a naturally degraded image (Stones). a) Reference image (exposure time 1/200 sec, aperture N=9), b) Naturally degraded image (exposure time 1/800 sec, aperture N=9) , c) Restored image using Wiener denoising, d) Restored image using the proposed method, e) Information removed by the Wiener method, f) Information removed by the proposed method.

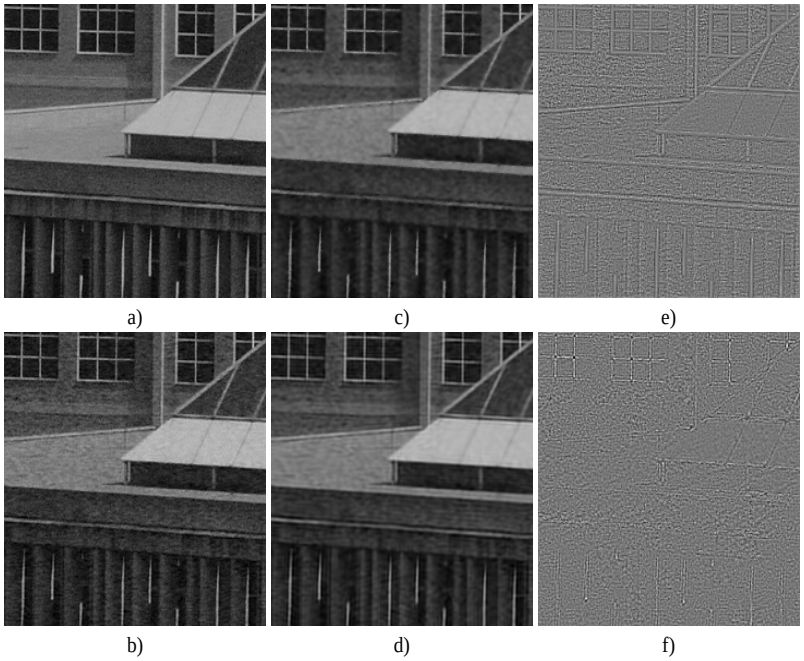


Fig. 15. Restoration results for a naturally degraded image (Building). a) Reference image (exposure time 1/1250 sec, aperture N=10), b) Naturally degraded image (exposure time 1/2000 sec, aperture N=10) , c) Restored image using Wiener denoising, d) Restored image using the proposed method, e) Information removed by the Wiener method, f) Information removed by the proposed method.

diagonalization of covariance and perceptual metric matrix in image coding," 2003, to appear in Pattern Recognition.

[33] J. Malo, "Normalized image representation for efficient coding," in *Proc. IEEE 37th Asilomar Conf. Signals, Systems and Comp.*, vol. 2, 2003, pp. 1408–1412.

[34] J. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proc.R.Soc.Lond. B*, vol. 265, pp. 359–366, 1998.

[35] F. Campbell and J. Robson, "Application of Fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, pp. 551–566, 1968.

[36] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM Journal of Sci. Comp.*, vol. 14, pp. 1487–1503, 1993.

[37] B. Girod, "What's wrong with mean squared error?" in *Digital Images and Human Vision*, A. B. Watson, Ed. MIT Press, 1993, pp. 207–220.

[38] P. Barten, "Evaluation of subjective image quality with the square root integral method," *JOSA A*, vol. 7, pp. 2024–2031, 1990.

[39] J. Lim, *Two-Dimensional Signal and Image Processing*. Prentice Hall, 1990.

[40] B. R. Hunt, "Digital image processing," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 693–708, 1975.

[41] J. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Pat. Anal. Mach. Intell.*, vol. 2, pp. 165–168, 1980.

[42] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 20, no. 6, pp. 1964–1977, 1999.

Juan Gutiérrez Juan Gutiérrez (1971) received the *Licenciado* degree in Physics (Electricity, Electronics and Computer Science) in 1995 from the Universitat de València.

He is with the Computer Science Department of the Universtitat de València, where he is an Assistant Professor. Currently, he is with the Visual Statistics Group (VI(S)TA) at the Universitat de València (http://www.uv.es/vista/vistavalencia). His current research interests include regularization theory, models of low-level human vision, and representation and analysis of images.

Jesús Malo Jesús Malo (1970) received the M.Sc. degree in Physics in 1995 and the Ph.D. degree in Physics in 1999 both from the Universitat de València.

He was the recipient of the Vistakon European Research Award in 1994. In 2000 and 2001 he worked as Fulbright Postdoc at the Vision Group of the NASA Ames Research Center (A.B. Watson), and at the Lab of Computational Vision of the Center for Neural Science, New York University (E.P. Simoncelli). Currently, he is with the Visual Statistics Group (VI(S)TA) at the Universitat de València (http://www.uv.es/vista/vistavalencia). He is member of the Asociación de Mujeres Investigadoras y Tecnólogas (AMIT).

He is interested in models of low-level human vision, their relations with information theory, and their applications to image processing and vision science experimentation. His interests also include (but are not limited to) Fourier, Matlab, modern art, independent movies, chamber music, Lou Reed, Belle and Sebastian, The Pixies, comics, la Bola de Cristal, and beauty in general...

Francesc J. Ferri Dr. Francesc J. Ferri (1964) received the *Licenciado* degree in Physics (Electricity, Electronics and Computer Science) in 1987 and the Ph.D. degree in Pattern Recognition in 1993 both from the Universitat de València.

He has been with the Computer Science and Electronics Department of the Universtitat de València since 1986; first as a research fellow and as a teacher of Computer Science and Pattern Recognition since 1988. He has been involved in a number of scientific and technical projects on Computer Vision and Pattern Recognition. He joined the Vision, Speech and Signal Processing Research group at the University of Surrey during a sabbatical in 1993 where he was working in Feature Selection and Statistical Pattern Recognition Methodology.

Dr. Ferri has authored or coauthored about 100 technical papers in international conferences and well-established journals in his fields of interest. He has also helped in refereeing in several of these journals and major conferences. His current research interests include Feature Selection, Nonparametric Classification Methods, Inductive Learning, Computational Geometry, Image Analysis and Motion Understanding.

He is a member of the ACM and the IAPR (AERFAI in Spain) where he has helped several times in organizing the biennal Spanish Conference on Pattern Recognition and Image Analysis.