1 November 2000

ELSEVIER

Optics Communications 185 (2000) 41–48

OPTICS COMMUNICATIONS

www.elsevier.com/locate/optcom

Improved rotation invariant pattern recognition using circular harmonics of binary gray level slices

Pascuala Garcia-Martinez [a,*], Henri H. Arsenault [b,1], Carlos Ferreira [a]

[a] *Dept. d'Òptica, Universitat de Valencia, Dr. Moliner, 50.46100 Burjassot, Spain*
[b] *COPL, Universite Laval. Ste-Foy, Qc, Canada G1K 7P4*

Received 30 May 2000; accepted 6 September 2000

Abstract

We introduce a new rotation invariant pattern recognition method based on nonlinear correlation. The images are decomposed into disjoint binary slices and then correlated using the common linear correlation. This operation is very discriminant even when the target is embedded in strong noise. We extend our sliced orthogonal nonlinear generalized correlation method to rotation invariant pattern recognition by combining the information of a circular harmonic (CH) of each binary slice of the reference object with binary slices of the target. In addition to improved discrimination capability, the method avoids the time-consuming process of finding proper centers for the CHs. Results are presented including the study of the stability of the correlation peaks in the presence of background noise and overlapping Gaussian noise.

Keywords: Rotation invariance; Pattern recognition; Morphological correlation; Joint transform correlator; Circular harmonic decomposition

1. Introduction

Although the human eye is fairly good at recognizing objects that are distorted by changes of orientation, scale and contrast, machines are better at some tasks than at others: for example machines are better at spotting targets buried in additive noise, but are generally not very good at taking scale changes into account.

The rotation problem has been extensively studied, and methods have been put forward for automatic recognition of targets under various orientations, especially for the in-plane rotation problem. The more general problem where the target can be rotated about all three axes is still unsolved except for special cases. In this paper we consider only in-plane rotations.

Some of the most effective methods for pattern recognition with in-plane rotations are based on the circular harmonic (CH) decomposition [1,2]. Although good results have been obtained, the constraint that only one circular harmonic component (CHC) from the target can be used per single filter has imposed performance limits. In order to solve this limitation and to improve the

* Corresponding author. Tel.: +34-96-3864717, ext. 23; fax: +34-96-3864715.

E-mail addresses: pascuala.garcia@uv.es (P. Garcia-Martinez), arseno@phy.ulaval.ca (H.H. Arsenault).

[1] Tel.: +1-418-656-2650; fax: +1-418-656-2623.

0030-4018/00/$ - see front matter
PII: S0030-4018(00)00979-2

discrimination capability (DC), many conventional filters have been combined with the CH decomposition [3–5]. However, some of those filters have poor noise robustness and one main drawback is that the performance depends on the choice of harmonic expansion center, which cannot be formulated in a systematic way.

In this paper we combine the use of multiple CHCs with a binary decomposition method. Moreover, the choice of the CH expansion centers can be done easily taking the mass center of the object. We are able to achieve rotation invariant pattern recognition under condition where other methods fail.

2. Binary decompositions and correlations

A two-dimensional (2D) image $g(x,y)$ with discrete gray levels can be decomposed into a sum of disjoint elementary images $e_m(g(x,y))$ satisfying the orthogonality property

$$\begin{aligned} &e_m(g(x,y))e_n(g(x,y)) = 0 \quad \text{if } m \neq n,\\ &e_m(g(x,y))e_n(g(x,y)) = 1 \quad \text{if } m = n. \end{aligned} \tag{1}$$

Each sub-image $\{e_m(g(x,y))\}$ represents a gray level slice of the object [6]. The sliced orthogonal nonlinear generalized (SONG) decomposition of $g(x,y)$ is [6]

$$g(x,y) = \sum_{i=0}^{Q-1} G_i e_i(g(x,y)), \tag{2}$$

where the coefficients G_i are weights, and where Q is the total number of gray levels in the image.

The elementary binary images have the property

$$e_i(g(x,y)) = \begin{cases} 1 & g(x,y) = i,\\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note that each object point has only one gray level, so each unshifted i-slice is disjoint and therefore orthogonal to all of the others. For the standard gray scale image representation, $G_i = i$.

The standard correlation between two objects $g(x,y)$ and $f(x,y)$ can be written

$$g(x,y) \otimes f(x,y) = \sum_{i=0}^{Q-1}\sum_{j=0}^{Q-1} G_i F_j e_i(g(x,y)) \otimes e_j(f(x,y)), \tag{4}$$

where $\otimes$ denotes the linear correlation and where the coefficients are equal to the gray levels, i.e. $G_i = i$ and $F_j = j$. Now the problem with this correlation is that it puts higher weights on brighter parts of the targets, but there is usually no reason why brighter parts of the targets should be more important than the others; that is why many pattern recognition techniques binarize both the reference and the target [7], thus giving equal weights to all the gray levels by setting all the weights equal to unity or some other constant value. This is convenient when the target may be segmented from the scene of which it is a part, but in highly cluttered or very noisy scenes, it is often not feasible to segment potential targets from the scene, nor to binarize the whole scene.

The coefficients G_i and F_j may be arranged into a matrix with rows and columns (i,j). We now generalized this matrix by replacing the product coefficients G_i and F_j by generalized weights W_{ij}. So the correlation expression (Eq. (4)) becomes

$$\mathrm{SONG}_{gf}(x,y) = \sum_{i=0}^{Q-1}\sum_{j=0}^{Q-1} W_{ij} e_i(g(x,y)) \otimes e_j(f(x,y)). \tag{5}$$

This is what we have called the SONG correlation [6]. Setting different values on the terms of the matrix $\mathbf{W}$ allows the matrix to represent various known correlation types such as standard matched filtering, binary filtering, morphological correlation [8], and so on; but it also allows us to define new kinds of correlation. The SONG correlation that we have proposed is obtained by setting

$$\begin{aligned} &W_{ij} = 0 \quad \text{for } i \neq j,\\ &W_{ij} = 1 \quad \text{for } i = j. \end{aligned} \tag{6}$$

In this case, the double sum reduces to the single sum and the correlation becomes

$$\Omega_{gf}(x,y) = \sum_{i=0}^{Q-1} e_i(g(x,y)) \otimes e_i(f(x,y)). \tag{7}$$

In this expression, only the gray levels in the two images having the same values are correlated together after having their values set equal to unity, so the correlation is a sum of correlations between binary images. This is the SONG correlation as used in our previous papers [6,9]. In the following, we shall consider for simplicity objects located such that their correlation peaks appear at the origin $(0,0)$; because the correlation operation is shift invariant, it is trivial to generalize to the case of targets located at point (x,y), and to the presence of multiple target, since the correlation is also additive.

Images are not usually digitized to fewer than 256 gray levels; but targets of interest in the scene usually cover a smaller range of levels, let us say 64 for purposes of discussion. It is clear that if 64 gray levels are maintained, expression shown in Eq. (7) will require summing 64 correlations. In fact such a large number of correlations is not required: it is simple matter to add neighboring gray levels together in order to reduce the number of gray levels to a more manageable number, say 4 or 8, which are the numbers used in our previous experimental results [6,8].

We have already shown some recognition results using the SONG method; we have also shown how the method can be implemented optically by means of a joint transform correlator (JTC) using additive techniques developed for morphological correlation [9]. So the SONG correlation can be considered either as a digital method in its own right or as an optical pattern recognition method. Another advantage of the SONG correlation is that if it is performed using binary information and a JTC, then it is convenient to implement it by a binary-valued SLM such as ferroelectric liquid crystals spatial light modulators (FLC SLMs) that can display 256×256 pixel^2 images at frame rates in excess of 1 kHz [10].

The SONG correlation, like the linear correlation, is shift invariant. However like the linear correlation, it is sensitive to object distortions such as rotations. In this paper we combine the SONG correlation with the CH decomposition in order to obtain rotation invariant pattern recognition.

3. Rotation invariant SONG correlation

The linear correlation between a scene $g(r,\theta)$ and a CHC of the reference object, $f_m(r,\theta)$ is rotation invariant [1,2]:

$$\gamma_m(r,\theta) = g(r,\theta) \otimes f_m(r,\theta), \tag{8}$$

where (r,θ) are the polar coordinates.

The circular harmonic filter (CHF) is a linear filter and is therefore shift invariant like the conventional matched filter. The CHF matches a single CHC of the target, resulting in rotation invariance [1].

The performance of the CHF involves both the order of the CHC and the expansion center chosen for the change from Cartesian coordinates to polar coordinates. Several methods have been introduced to select the appropriate order and center, called the proper center [11,12].

We apply the CH decomposition to each binary elementary function $(e_i(f(x,y))$ of Eq. (7), which yields

$$\begin{aligned}\Omega_{gf}(r,\theta) &= \sum_{i=0}^{Q-1} e_i(g(r,\theta)) \otimes e_i(f(r,\theta)) \\ &= \sum_{i=0}^{Q-1} e_i(g(r,\theta)) \otimes \sum_{m=-\infty}^{\infty} e_{i_m}(f(r,\theta)) \\ &= \sum_{m=-\infty}^{\infty} \sum_{i=0}^{Q-1} e_i(g(r,\theta)) \otimes e_{i_m}(f(r,\theta)).\end{aligned} \tag{9}$$

Rotation invariant pattern recognition is achieved by using only one component of the CH decomposition and then performing the linear correlation. If only one CHC is used for each i-correlation in Eq. (9), the SONG correlation will allow the detection of a target for any angular orientation. So using only one m-order CHC in Eq. (9), we define a rotation invariant SONG (RISONG) correlation as

$$\Omega_{gf}^{m}(r,\theta) = \sum_{i=0}^{Q-1} e_i(g(r,\theta)) \otimes e_{i_m}(f(r,\theta)). \tag{10}$$

The correlation of Eq. (10) is also a nonlinear correlation. The difference between the RISONG and the SONG is that only part of the binary

elementary slice $(e_i(f(x,y))$ is used in the correlation.

The RISONG can be implemented optically using a JTC in a manner similar to that used for the optical implementation of another rotation invariant nonlinear correlation called the rotation invariant morphological correlation (RIMC) [13]. The different binary slices with the CHC of the reference object and the input scene are displayed in a SLM. All the joint power spectra (JPSs) are obtained and added. This final JPS distribution is displayed again in the SLM, then a second Fourier transformation will yield the RISONG correlation, see Eq. (10). The main issue in this optical implementation is to display the complex CHC in the JTC. There are methods to display complex functions in a SLM of the JTC [14,15]. For the optical implementation of the RIMC [13], we used the real $m = 0$ CHC of each binary slice reference object. Although the zero-order CH order has poor information about the object, the result showed that the RIMC was satisfactory. However, if the objects are quite similar, the method can give a false detection. In order to obtain higher discrimination, we propose the RISONG correlation.

4. Results

In this section we present some computer results to determine the performance of the method for rotation invariant pattern recognition and to compare the performance with other linear CH pattern recognition methods.

Fig. 1 shows an input scene that is made up of different vehicles: two rotated versions of the reference objects (0° and 90°) are marked with arrows. The image has 16 gray levels.

The choice of the order and of the expansion center are important for the RISONG case as it is for the CHC in the linear case, but as discussed below, the choice of expansion center for the RISONG is much simpler than for the latter.

The function $f_m(x,y)$ in Eq. (8) takes into account the choice of the CH order and of the expansion center as discussed elsewhere [11,12]. For the RISONG, the CHCs of the elementary binary *i*-slices of the reference object are required. We

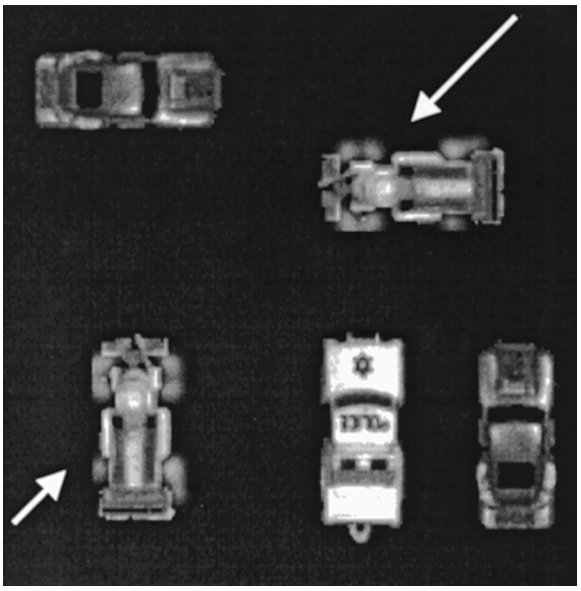

Fig. 1. Input scene with two reference objects rotated (0° and 90°).

chose the same expansion order as for the linear case, because that parameter is strongly dependent on the geometry of the reference object as well as on the shape of each binary slice which in turn is strongly dependent on the shape of the reference object. For the objects shown in Fig. 1 we used $m = 5$ CHFs.

For the RISONG expansion center, we simply chose the mass center of the reference object for all the *i*-CHC binary slices, because the gray scale distribution is related to that mass center of the object, and it is important for all the CHs of the slices to have the same expansion center. So for the RISONG correlation the choice of expansion center is trivial because the mass center is easy to calculate, and we avoid the laborious task of finding the proper expansion center, which is a prerequisite to obtain good performance in the case of a single CHC filter.

We define the DC as

$$\mathrm{DC} = 1 - \frac{\mathrm{CrossCorr}}{\mathrm{AutoCorr}}, \tag{11}$$

where the AutoCorr and the CrossCorr are the auto-correlation peak value and the cross-correlation peak value. Taking into account that we

used many objects in the input scene, we will obtain different cross-correlation peak values. The DC is calculated using the highest cross-correlation peak value.

A high value of DC means that the value of the cross-correlation is low compared to the auto-correlation, which implies that good discrimination and good noise robustness are achieved. On the other hand, a low value of the ratio means that the energy of the cross-correlation has almost the same value as that of the auto-correlation; so the higher the better.

The linear correlation for an $m = 5$ CHF is shown in Fig. 2(a). Note that this filter is unable to detect the reference object, and a false alarm appears. On the other hand, if we use the same input scene with the RISONG, we obtain the correlation output shown in Fig. 2(b). The DC for this case is 0.93. We used the $m = 5$ CHC order for each binary slice.

The correlation peaks in Fig. 2(b) are much sharper. Other linear filters that are more discriminant than the matched filter, for example the CH phase only filter (POF) or CH inverse filter (IF) do not yield better results. This superior performance of the RISONG method can be understood from the orthogonality property at the origin of the binary elementary functions: although the different CHCs of the elementary binary i-slices $(e_{i_m}(f(r,\theta)))$ are not disjoint or orthogonal, they do not correlate well with each other because they correspond to slices that are very different and that are disjoint (in the case of a true target) or quasi-disjoint (in the case of a false target). These computer results demonstrate the improved performance of the RISONG over the CHC linear correlation.

Although the performance of the SONG and of the RISONG correlation is very good, both operations are quite sensitive with respect to change of illumination or changes of gray level distribution. A complete solution of this problem is beyond the scope of this paper, but we will introduce solutions for a few cases. One such case is where the illumination source is known, or may be determined for example by knowledge of the illumination source or by examination of a known part of the scene; it is rather trivial in such cases to normalize the whole scene and to adjust either or both the reference and the target scene so that they have the same number of gray levels, after which the SONG correlation and the RISONG correlation may be applied as described above.

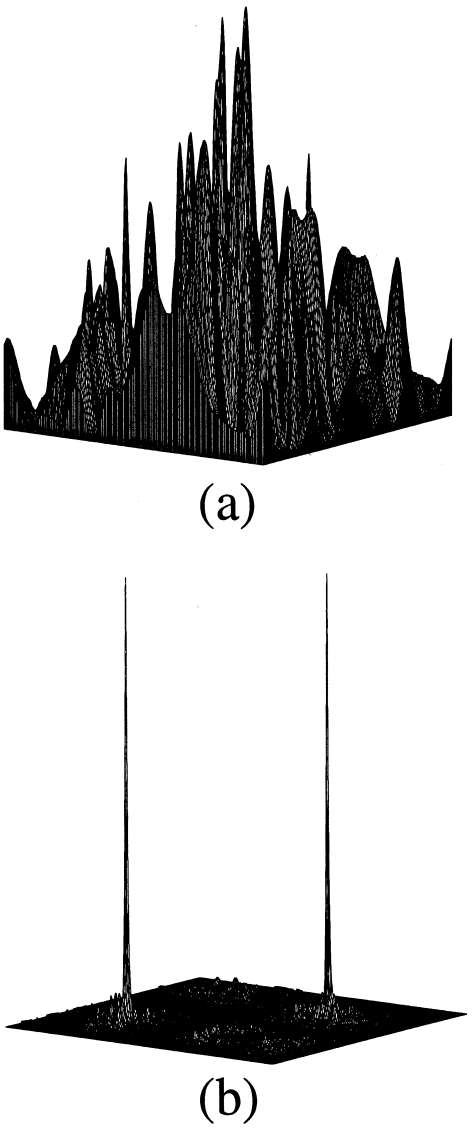

Fig. 2. (a) 3D plot of the CH linear correlation for Fig. 1. (b) 3D plot of the RISONG correlation for Fig. 1.

If the illumination parameter is unknown, then the solution can be to quantize the image after acquiring it with a camera. Real images have many gray levels and those gray levels are not uniform

distributions. In such cases, a variation in the gray scale will not dramatically affect the correlation if we quantize using a small number of quantization levels. We are aware of the importance of quantization for both the SONG and the RISONG, this is why we are studying it carefully; but we feel that this illumination study is beyond the scope of this rotation invariant pattern recognition paper, and that there are many cases where illumination changes is not a problem or can be dealt with easily. This is especially the case when potential targets may be segmented from the scene.

Here we will consider correlated disjoint noise and additive Gaussian noise.

5. The effect of noise on performance

Correlated Gaussian disjoint noise and additive Gaussian noise are two of the most important types of noise that degrade images. We first consider the stability of the DC of the RISONG correlation for the detection of targets in the presence of disjoint correlated noise. Fig. 3 shows the scene of Fig. 1 corrupted by correlated Gaussian disjoint noise. The background is a Gaussian noise distribution with a mean of zero and various values of the standard deviation σ; this parameter is a measure of the energy of the noise pattern, where a greater value of σ represents a higher level of noise.

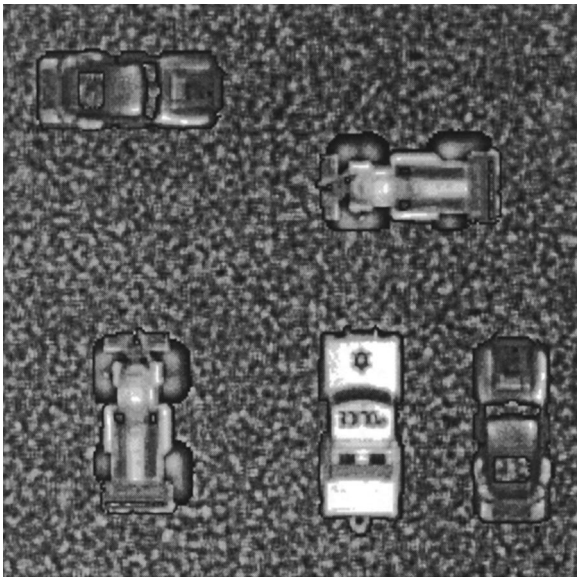

Fig. 3. Input scene corrupted with correlated Gaussian disjoint noise.

Table 1 shows the performance of the RISONG DC for various values of the nonoverlapping noise. Note that the RISONG is stable over a wide range of noise levels. Standard detection methods like the CH linear correlation, the CH POF or the CH IF are unable to detect the targets over the whole range of noise. All of them yield false alarms.

We also studied the performance of the RISONG for overlapping Gaussian Noise. In Fig. 4 is shown the same scene of Fig. 1 corrupted by additive Gaussian Noise. The third column of Table 1 gives the DC parameters for various values of the noise for the RISONG, and the other columns correspond to the CH linear and the CH POF. Both of the latter filters detected the brighter car, yielding a false detection (DC = 0).

Fig. 5 shows the RISONG correlation for $\sigma = 1.9$; although the two peaks are slightly different, they correspond to the two rotated versions of the reference object, and there is plenty of leeway for a threshold detection.

6. Conclusion

We have introduced a rotation invariant pattern recognition method that yields excellent results for cases where targets are difficult to recognize and that yields stable correlation peaks over a wide range of high noise levels where other methods fail. The method uses multiple filters, so the price paid is an increase in processing time. However this increase is more than compensated by the removal of the need to calculate a proper center for CHFs, a very time-consuming operation and by the improved performance.

We do not claim that this method is superior to every existing method, because there is no method under the sun that can make that claim; some methods perform better in some circumstances, and less well in other cases. No doubt that will be the case for this method, even after we find a better

Table 1
DC of several pattern recognition operations and the RISONG when different nonoverlapping and overlapping Gaussian noise degrees are considered

Gaussian noise standard deviation (σ)	RISONG		CH matched filter (overlapping noise)	CH POF (overlapping noise)
	Nonoverlapping noise	Overlapping noise		
0	0.95	0.93	0	0
0.25	0.93	0.90 ± 0.02	0	0
0.5	0.95	0.92 ± 0.02	0	0
0.75	0.93	0.91 ± 0.02	0	0
1	0.93	0.90 ± 0.02	0	0
1.5	0.93	0.90 ± 0.02	0	0
1.9	0.93	0.90 ± 0.02	0	0

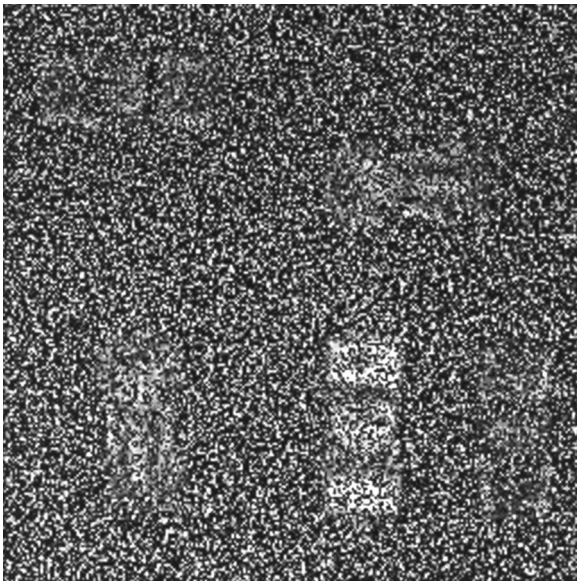

Fig. 4. Input scene of Fig. 1 with overlapping Gaussian noise ($\sigma = 1.9$).

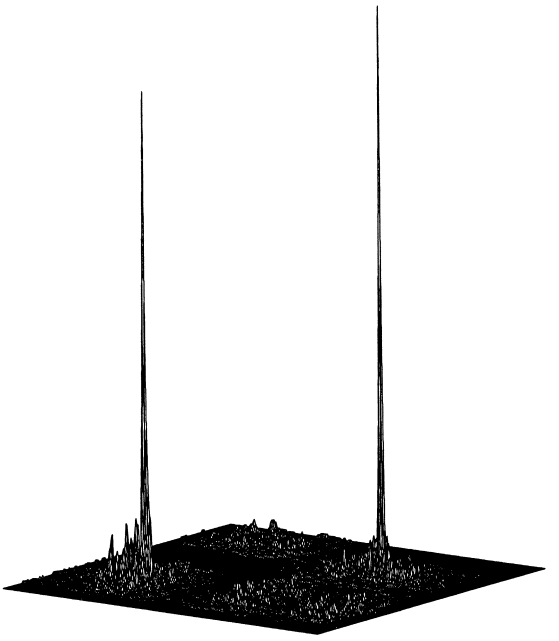

Fig. 5. 3D plot of the RISONG correlation for Fig. 4.

solution to the intensity invariance problem. For doubtful readers, we are well on the way to finding such a solution, but that is beyond the scope of this paper.

Acknowledgements

This work has been supported by the Spanish DGES, Dirección General de Enseñanza Superior, project PB96-1134-C02-02 and by a grant from the natural Sciences and Engineering Research Council of Canada.

References

[1] Y.N. Hsu, H.H. Arsenault, G. April, Appl. Opt. 21 (1982) 4012.
[2] Y.N. Hsu, H.H. Arsenault, Appl. Opt. 21 (1982) 4016.
[3] H.F. Yau, C.C. Chang, Appl. Opt. 28 (1989) 2070.
[4] Ph. Refregier, Opt. Commun. 86 (1991) 113.

[5] G. Ravichandran, D.P. Casasent, Opt. Eng. 30 (1991) 1601.
[6] P. Garcia-Martinez, H.H. Arsenault, Opt. Commun. 174 (2000) 503.
[7] W.C. Hasenplaugh, M.A. Neifeld, Opt. Eng. 38 (1999) 1907.
[8] P. Garcia-Martinez, D. Mas, J. Garcia, C. Ferreira, Appl. Opt. 37 (1998) 2112.
[9] P. Garcia-Martinez, H.H. Arsenault, S. Roy, Opt. Commun. 173 (2000) 185.
[10] H. Sjöberg, B. Noharet, R. Hey, Opt. Eng. 37 (1998) 1316.
[11] Y. Sheng, H.H. Arsenault, J. Opt. Soc. Am. A 4 (1987) 1793.
[12] P. Garcia-Martinez, J. Garcia, C. Ferreira, Opt. Commun. 117 (1995) 399.
[13] P. Garcia-Martinez, C. Ferreira, J. Garcia, H.H. Arsenault, Appl. Opt. 39 (2000) 776.
[14] R. Piestun, J. Rosen, J. Shamir, Appl. Opt. 33 (1994) 4398.
[15] P. Birch, R. Young, C. Chatwin, M. Farsari, D. Budgett, J. Richardson, Opt. Commun. 175 (2000) 347.