VNIVERSITAT DE VALÈNCIA

CWI

Dynamic Probabilistic Multimedia Retrieval Model

Tzvetanka I. Ianeva

Introduction

Analysis of a video as a single entity by modeling one-second video sequence around the keyframe

Gaussian Mixture Model in DCT-space-time domain

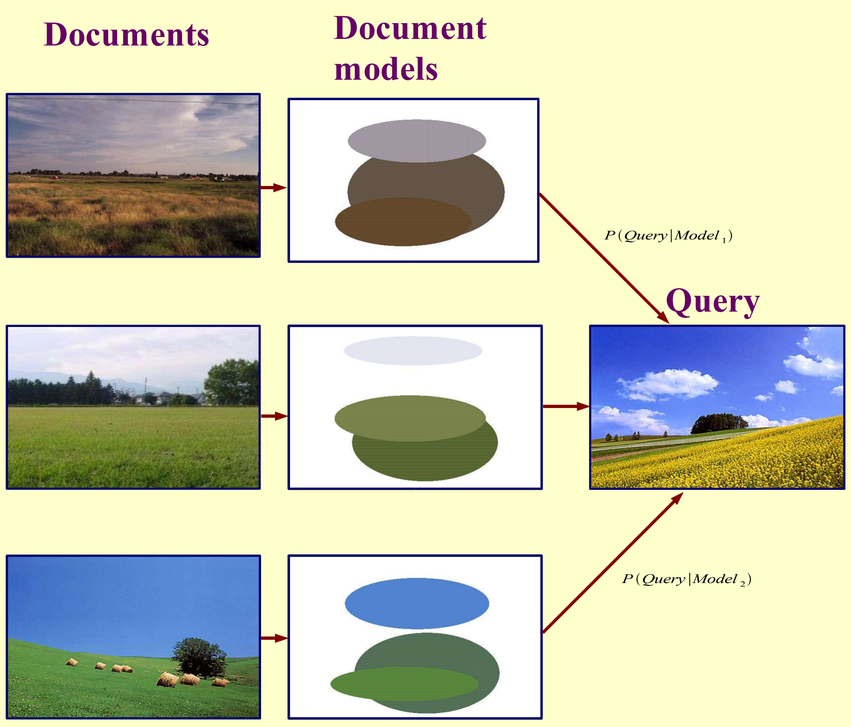
Extension of the Static Probabilistic Retrieval model
Thijs Westerveld, Arjen P. de Vries: Experimental result analysis for a generative probabilistic image retrieval model. SIGIR2003: 135-142

Pursue *video* retrieval instead of keyframe retrieval

Static Retrieval Model

Given query image $\boldsymbol{x}$ consisting of N samples $(\boldsymbol{x} = x_1, x_2, ..., x_N)$, collection image (document) w_i is compared to $\boldsymbol{x}$ by computing its ability ***(RSV)*** to explain the samples $\boldsymbol{x}$:

$$RSV(w_i) = \frac{1}{N}\sum_{j=1}^{N} \log[kP(x_j|w_i) + (1-k)P(x_j)]$$



Building Dynamic GMMs

The Dynamic GMM captures the disappearing of the grass in the video sequence. The corresponding component (green dynamic space-time blob) dissapears at about t=0.4

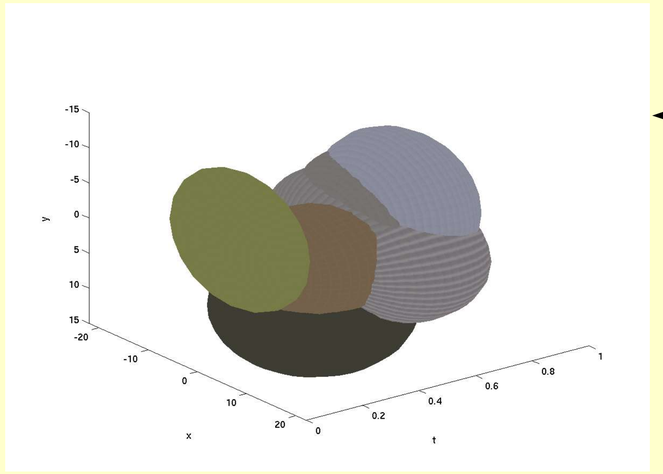
This effect is impossible to be captured from the static model where time is not taking into account

split color channels

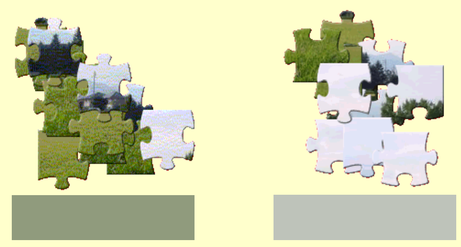
DCT coefficients | position | time

642,12	17	-341,77	-23,55	-9,93	5,11	1,03	-1,26	4	-23,55	952,87	989,62	1	1	0,04
596	-7,53	-349,53	-8,21	2,43	-3,19	6,26	2,78	6,24	-5,42	954,75	993,25	1	2	0,04
595,75	0,17	-346,96	-12,21	-6,85	6,56	0,94	-0,58	10,08	-9,33	956,75	989,5	1	3	0,04
591,62	11,87	-332,65	-28,67	-2,86	-5,85	0,07	9,77	-7,5	-4,94	938,75	989,25	1	4	0,04
589,5	-5,8	-333,71	-24	13,61	4,89	-8,21	-1,84	-12,53	-6,82	956,75	993,75	1	5	0,04
607	2,65	-356,02	-14,96	-1,22	2	0,49	3,74	-0,67	-9,91	956,75	983,25	1	6	0,04
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
128,12	0,17	-0,03	-0,16	-0,05	0,16	0,15	-0,05	-0,23	0,1	1024	1024	22	15	1

EM algorithm

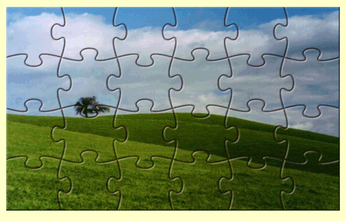


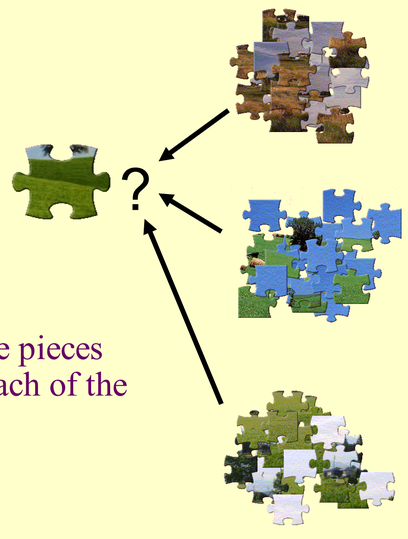
Like sorting the pieces of a jigsaw
- Assign pieces to random component
- Compute means for each class
- Re-assign pieces



Probabilistic Image Retrieval

Example
Query:





What is the probability that the pieces of this query are drawn from each of the documents models?

Model [Vasconcelos 2000]
Gaussian Mixture Model (GMM)
each sample from an image is generated by 1 of C components

$$P(x|w_i) = \sum_{c=1}^{N_C} P(C_{i,c})\,G(x, \mu_{i,c}, \Sigma_{i,c})$$

$$G(x, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

Conclusions/Future Work

Conclusions

- appropriate keyframe less critical
- can model (dis)appearance of objects
- modeling sequence around the keyframe better than taking not connected frames from the full shot
- combining textual and dynamic visual models better than ASR only
- best results with multiple examples round-robin merged, topical filters and non-visual modalities

Future Work

- more data needs more computation effort – optimizations ?
- integration of audio
- modeling of motion and texture over time