MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Factorized Local Appearance Models

Baback Moghaddam, Xiang Zhou

TR2002-51 December 2002

Abstract

We propose a novel local appearance modeling method for object detection and recognition in cluttered scenes. The approach is based on the joint distribution of local feature vectors at multiple salient points and factorization with Independent Component Analysis (ICA). The resulting non-parametric densities are simple multiplicative histograms. This leads to computationally tractable joint probability densities which can model high-order dependencies. Testing and evaluation shows that the factorized density model with spatial encoding improves modeling accuracy and outperforms global appearance models in image/object retrieval. Furthermore, experiments in detection of substantially occluded objects in cluttered scenes have demonstrated promising results.

International Conference on Pattern Recognition (ICPR02)

Published in: *International Conference on Pattern Recognition* (ICPR'02), Canada, August 2002.

Factorized Local Appearance Models

Baback Moghaddam
Mitsubishi Electric Research Laboratory
Cambridge, MA 02139 USA

Xiang Zhou
Beckman Institute, University of Illinois
Urbana-Champaign, IL 61801 USA

Abstract

We propose a novel local appearance modeling method for object detection and recognition in cluttered scenes. The approach is based on the joint distribution of local feature vectors at multiple salient points and factorization with Independent Component Analysis (ICA). The resulting non-parametric densities are simple multiplicative histograms. This leads to computationally tractable joint probability densities which can model high-order dependencies. Testing and evaluation shows that the factorized density model with spatial encoding improves modeling accuracy and outperforms global appearance models in image/object retrieval. Furthermore, experiments in detection of substantially occluded objects in cluttered scenes have demonstrated promising results

.

1. INTRODUCTION

For appearance based object modeling in images, the choice of method is usually a trade-off determined by the nature of the application or the availability of computational resources. Existing object representation schemes provide models either for global features[16], or for local features and their spatial relationships [13][1][15][7]. With increased complexity, the latter provides higher modeling power and accuracy.

Among various local appearance and structure models, there are those that assume rigidity of appearance and viewing angle, thus adopting more explicit models [15][13][11]; while others employ stochastic models and use probabilistic distance and matching metrics [7][10][1].

We construct a probabilistic appearance model with an emphasis on the representation of non-rigid and approximate local image structures. We use joint histograms on k-tuples (k salient points) to enhance the modeling power for local dependency, while reducing the complexity by histogram factorization along the feature components. Unlike [15], in which sub-region dependency is intentionally ignored for simplicity, we explicitly model the dependency by joint histograms. Although, the gain in modeling power of joint densities increases the computational complexity, we propose histogram factorization based on an Independent Component Analysis (ICA) [2] to dramatically reduce the histogram dimensionality, thus reducing the computation to a level that can be easily handled by today's personal computers.

In this paper, we will focus our attention on the modeling of images and objects through the use of joint histograms. Figure 1 provides an overview diagram of our histogram-based image and object model. More detailed description is given in Section 2. This model has been applied toward image retrieval and object detection in cluttered scenes (Section 4) with promising results.

2. METHODOLOGY

We propose multi-dimensional histograms as a non-parametric approximation of the joint distribution of image features at multiple image locations. Let i be the index for elementary feature components in an image, which can be pixels, corner/interest points [4][6], blocks, or regions in an image. Let $\boldsymbol{x_i}$ denote the feature vector of dimension n at location i. $\boldsymbol{x_i}$ can be as simple as {R, G, B} components at each pixel location or some invariant feature vectors extracted at corner or interest points [9][13][14] or even transform domain coefficients at an image block, or any other local/ regional feature vectors.

For model-based object recognition, we use the *a posteriori* probability

$$\max_l P(M_l \mid T) \qquad (1)$$

where M_l is the object model and $T = \{\boldsymbol{x}_i\}$ represents the features found in the test image. Equivalently, by assuming equal priors, classification/detection will be based on maximum likelihood testing:

$$\max_l P(T \mid M_l) \qquad (2)$$

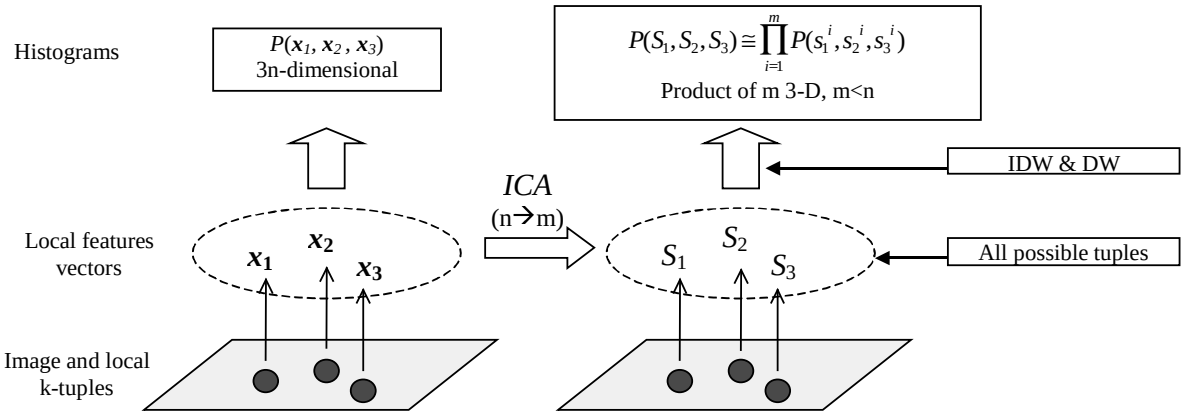

Figure 1 : Image local appearance modeling by joint histograms

For the class-conditional density in Equation (2), it is intractable to model dependencies among all $\boldsymbol{x}_i$'s (even if correspondence is solved), yet to completely ignore these dependencies will severely limit our modeling power. Objects frequently distinguish themselves not by individual regions (or parts), but by the relative location and appearance of these regions. A tractable compromise between these two modeling extremes (which also does not require correspondence) is to model the joint density of all *k*-tuples of $\boldsymbol{x}_i$'s in *T*.

2.1 Joint distribution of *k*-tuples

Instead of modeling the total joint likelihood of all $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_I$, which is an $(I \times n)$-dimensional distribution, we model of the distribution of all *k*-tuples as an approximation:

$$P(\{(x_{i_1}, x_{i_2}, \ldots, x_{i_k})\} \mid M_l) \tag{3}$$

This becomes a $(k \times n)$-dimensional distribution, which is still intractable. For example, for 20 histogram bins along each dimension, we have $20^{(k \times n)}$ bins. Therefore, we factorize this distribution into a product of low-dimensional distributions. We achieve this factorization by transforming $\boldsymbol{x}$ into a new feature vector $\boldsymbol{S}$ whose components are (mostly) independent. This is where independent component analysis (ICA) comes in.

2.2 Histogram factorization based on ICA

ICA originated in the context of blind source separation [8][2] to separate "independent causes" of a complex signal or mixture. It is usually implemented by pushing the vector components away from Gaussianity by minimizing high-order statistics such as the 4th order cross-cumulants. ICA is in general not perfect therefore the IC's obtained are not guaranteed to be completely independent.

By applying ICA to $\{\boldsymbol{x}_i\}$, we obtain the linear mapping

$$x \approx AS \tag{4}$$

and

$$P(\{(S_{i_1}, S_{i_2}, \ldots, S_{i_k})\} \mid M_l) \approx \prod_{j=1}^{m} P(\{(s_{i_1}^j, s_{i_2}^j, \ldots, s_{i_k}^j)\} \mid M_l) \tag{5}$$

where *A* is a n-by-m matrix and S_i is the "source signal" at location *i* with nearly independent components. The original high-dimensional distribution is now factorized into a product of *m* *k*-dimensional distributions, with only small distortions expected. We note that this differs from so-called "naïve Bayes" where the distribution of feature vectors is assumed to be factorizable into 1-D distributions for each component. Without ICA the model suffers since in general these components are almost certainly statistically *dependent*.

After factorization, each of the factored distributions becomes manageable if *k* is small, e.g., *k* = 2 or 3. Moreover, matching can now be performed individually on these low-dimensional distributions and the scores are additively combined to form an overall score.

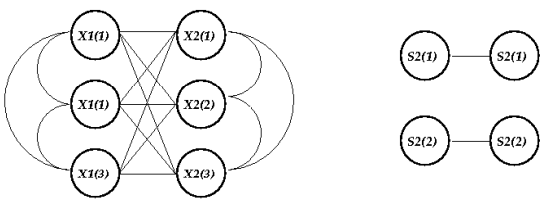

Figure 2 Graphical Models: (a) fully-connected graph denoting no independence assumptions (b) the ICA-factorized model with pair-wise only dependencies.

Figure 2 is a graphical model showing the dependencies between a pair of 3-dimensional feature vectors $\boldsymbol{x}_1, \boldsymbol{x}_2$. The joint distribution over all nodes is 6-dimensional and all nodes are (potentially) interdependent. The basic approach towards obtaining a tractable distribution is to remove intra-component dependencies (vertical and diagonal links) leaving only inter-component dependencies (horizontal links). Simultaneously, we seek to reduce the number of components from n=3 to m=2 "sources". Ideally, a perfect ICA transform results in the graphical model shown in the right diagram where the pair $\boldsymbol{S}_1, \boldsymbol{S}_2$ only have pair-wise inter-component dependencies. Therefore, the resulting factorization can be simply modeled by only two 2-D histograms in this case.[1]

3. EXPERIMENTS

We have tested the new model in the applications of object detection and image retrieval. For object detection we used synthetic "cluttered" images that are actually a collage of multiple object images.

3.1 Object Detection

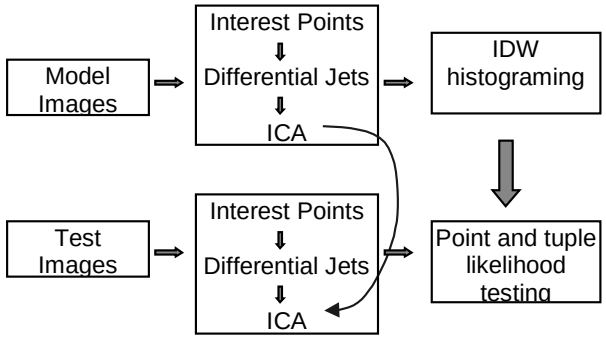

Figure 3. Diagram for object detection and localization (arrow indicates the same ICA basis is used)

First, tests on object detection in cluttered scenes were conducted. Figure 3 shows the flow diagram for this task. In Figure 3, note that we use the ICA mixing matrix *A* of the model images on the test images for direct computation of their IC's. This is based on the intuition that if the test image is cluttered, its own mixing matrix will not agree with that of the model. This in turn can distort a potential candidate's ICA components.

[1] We should note that in practice with an approximate ICA transform, the diagonal links of the original model are less likely to be removed than the vertical ones.

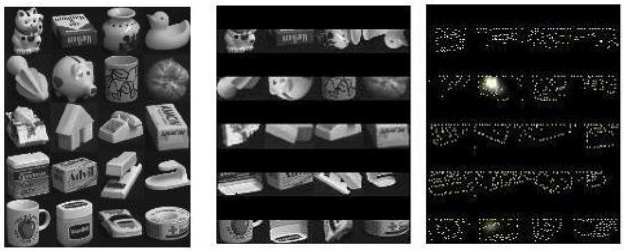

Figure 4 Synthetic "cluttered" scene and a detection example. (a) The synthetic test image of 20 objects from COIL; (b) The rotated and occluded version of (a); (c) The likelihood map for detecting "piggy bank" in (b). The white dots are the interest points. High-likelihood points are highlighted.

In our experiments we used a Harris operator [6][14] to detect interest points and computed the first 9 differential invariant jets [9] at each point as the corresponding feature vector $\boldsymbol{x}$. We must emphasize however that our methodology is not restricted to differential invariant jets and can in principal be used for any local set of features; for example, color, curvature, texture, edge-density, texture moments, etc. An ICA was then performed to get *m* independent components. We used $k = 2$, resulting in a set of 2-D histograms which were used to model 2-tuple joint component densities.

Test images were constructed using 20 objects from the Columbia Object Image Library (COIL) [12] (Figure 4(a)). To test the invariance properties, each of the objects is transformed by 3-D pose change, a planar rotation, followed by 50% occlusion (Figure 4(b)). Figure 4(c) shows the raw output for "piggy bank" detection on (b) where high-likelihood points have higher intensity.

The effectiveness of ICA was evaluated by comparing 1 through 9 IC's with the original 9 jets as the feature vector. For the original 9 jets, the histogram factorization along feature components is no longer valid (leading to "naïve Bayes") since the independence assumption on the jets does not hold in general.

Detection performance was measured by the average rank of the *accumulated regional likelihood* for the model object (the ground truth object location was used). Figure 5 depicts the clear improvement introduced by ICA. By using 3 IC's the system achieved 100% "first guess" detection (average rank = 1) on Figure 4(a), and an averaged rank of 1.2 for Figure 4(b), in which each object is *rotated* and *occluded*. For *pose change* of 10^o, the average rank is 2.75, which means that an object is detected on average within the first 3 locations checked.

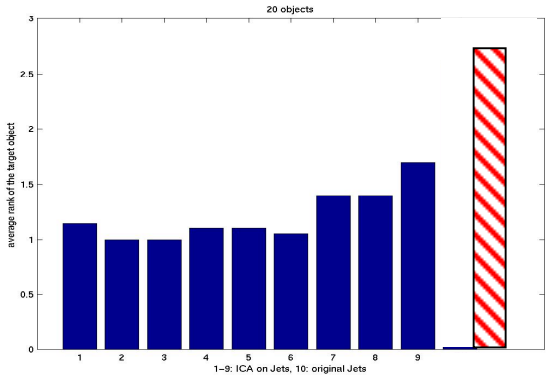

Figure 5. The average detection rank of the target object using ICs (m = 1, 2, … 9) vs. original 9-dimensional jets (shown as the rightmost bar). Dataset: COIL; 20 objects

3.2 Image Retrieval

We also tested the new model for image retrieval on two data sets: a subset of 20 objects from COIL, with 5 images at adjacent poses for each object; the other is a subset of 70 images from COREL photo images, with 7 classes and 10 images from each class.

Each image in the set is used as the query and its histogram as the model for comparison with all other images. The averaged hit-rate for the first candidate returned (which is also the nearest neighbor classification accuracy) is used as the performance measure. We compared multiple distance metrics, including Kullback-Leibler (KL) distance [3], Chi-squared distance [5], and Histogram Intersection (HI) [16]. From our experiments, we found out that these metrics yield statistically comparable results.

Table 1 Performance of the factorized local apperance model against a global feature method for image retrieval

Data set	COIL	COREL
Global Texture/Structure	96%	91%
Factorized Multi-Jets	**97**%	**96**%

To compare our local method to a more traditional global one, we combined wavelet moments as texture features and water-filling features as structural features, and used Euclidean distance measure. The comparison is listed in Table 1. Here Histogram Intersection was used as our distance measure. We see that the new representation yields comparable, if not better, retrieval results on both data sets.

5 DISCUSSION

A novel probabilistic modeling scheme was proposed based on factorization of high-dimensional distributions of local image features. We argued in favor of the *k*-tuple histograming scheme for the purpose of capturing local spatial dependencies. A distinct advantage of the proposed method is the flexibility in modeling spatial relationships (by varying *k*). Experiments have yielded promising results in image retrieval as well as in robust object localization in cluttered scenes. In the future, we plan to explore a more explicit way to incorporate spatial adjacency into the factorized local appearance model via graph matching or with local shape descriptors.

REFERENCES

[1] P. Chang, J. Krumm, “Object recognition with color cooccurrence histograms”, CVPR’99, Colorado, June. 1999

[2] P. Comon, “Independent component analysis – a new concept?” Signal Processing 36:287-314, 1994

[3] T. Cover and J. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., New York. 1991

[4] R. Deriche and G. Giraudon, “ A computational approach for corner and vertex detection”, Int’l Journal of Computer Vision, vol. 10, no. 2, pp. 101-124, 1993

[5] K. Fukunaga, “Introduction to Statistical Pattern Recognition,” Academic Press, 1971

[6] C. Harris and M. Stephens, “A combined corner and edge detector”, in Alvey Vision Conf., 1988, pp. 147-151

[7] J. Huang, S. R. Kumar, M. Mitra, W.-J., Zhu, and R. Zabih, “Image indexing using color correlograms,” IEEE conf. on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 1997

[8] C. Jutten, and J. Herault, “ Blind separation of sources,” Signal Processing, 24:1-10, 1991

[9] J. J. Koenderink, and A. J. van Doorn, “Representation of local geometry in the visual system,” Biological Cybernetics, vol. 55, pp. 367-375, 1987.

[10] B. Moghaddam, H. Biermann, D. Margaritis, "Regions-of-Interest and Spatial Layout in Content-based Image Retrieval," in European Workshop on Content-Based Multimedia Indexing, CBMI'99, France, Oct. 1999.

[11] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," Pattern Analysis and Machine Intelligence, PAMI-19(7), pp. 696-710, Jul. 1997

[12] S. A. Nene, S. K. Nayar and H. Murase. Columbia Object Image Library: COIL-100. Technical Report CUCS-006-96, Department of Computer Science, Columbia University, February 1996

[13] C. Schmid and R. Mohr, “Local grayvalue invariants for image retrieval”, PAMI, 19(5): 530-534, 1997

[14] C. Schmid, R. Mohr, and C. Bauckhage, “Comparing and evaluating interest points”, Proc. ICCV, 1998

[15] H. Schneiderman, T. Kanade "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object recognition, CVPR’98, pp. 45-51. 1998. Santa Barbara, CA.

[16] M. J. Swain and D. H. Ballard, “Color Indexing,” Int’l Journal of Computer Vision, vol. 7, pp. 11-32, 1991.