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

Factorization for Probabilistic Local Appearance Models

Baback Moghaddam
Xiang Zhou

TR2002-50 June 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. Furthermore, we propose a distance-sensitive histograming technique for capturing spatial dependencies which are otherwise lost in the joint feature distributions. The advantages over existing techniques include the ability to model non-rigid objects and the flexibility in modeling spatial or structural relationships between object parts. 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.

Int'l Conf. on Signal Processing, Pattern Recognition & Applications (SPPRA'02)
Crete, Greece, June 2002

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Copyright © Mitsubishi Electric Research Laboratories, Inc., 2002
201 Broadway, Cambridge, Massachusetts 02139

Published in: *Int'l Conf. on Signal Processing, Pattern Recognition & Applications* (SPPRA'02)

Factorization for Probabilistic Local Appearance Models

Baback Moghaddam and Xiang Zhou
Mitsubishi Electric Research Laboratory
University of Illinois at Urbana-Champaign

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. Furthermore, we propose a distance-sensitive histograming technique for capturing spatial dependencies which are otherwise lost in the joint feature distributions. The advantages over existing techniques include the ability to model non-rigid objects and the flexibility in modeling spatial or structural relationships between object parts. 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 can increase 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.

For modeling local structures, we use distance-sensitive histograming technique. In [7] and [1], the distance information is explicitly captured into the histogram bins. We argue in favor of collapsing the distance axis and instead using weighted histogram bin counts which are distance–dependent (proportional or inverse-proportional) . For example, for articulated and non-rigid object, any constraint on the structure or distance between distant points/regions can be misleading. In this case, inverse-distance-weighted histograming is the preferred method. The choice of which weighting scheme to use is application-dependent.

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 joint 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.

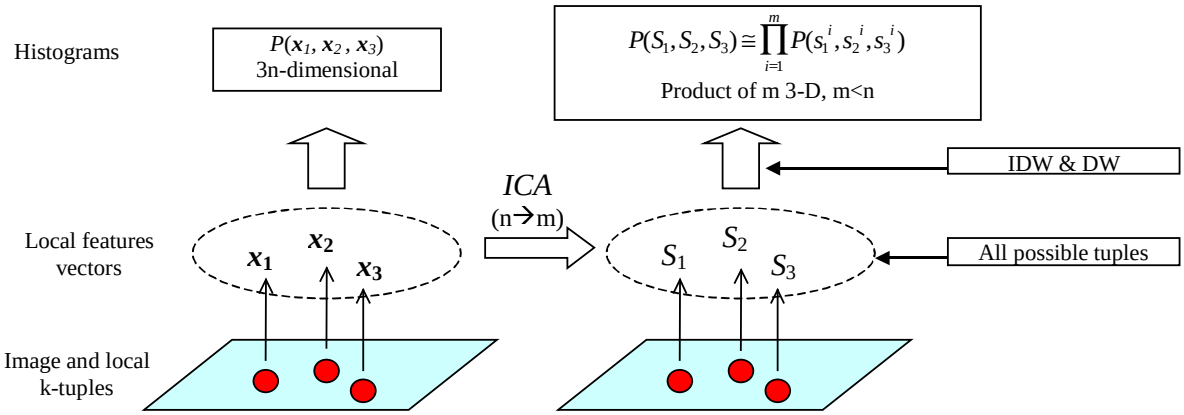


Figure 1 : Image local appearance modeling by joint histograms

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

$$\max_l P(M_l \mid T) \tag{1}$$

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:

$$\max_l P(T \mid M_l) \tag{2}$$

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

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

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) is used to effectively factorize the joint probability density function into a product of 1D marginal densities which are captured by means of histograms (as opposed to parametric models) for greatest flexibility in modeling.

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

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

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 formulation differs from the so-called "naïve Bayes" approach whereby the distribution of individual feature vector components is assumed to be already independent and hence factorizable into 1-D distributions. Without ICA factorization the resulting density model ultimately suffers since in general the components of local image features are almost certainly statistically *dependent*.

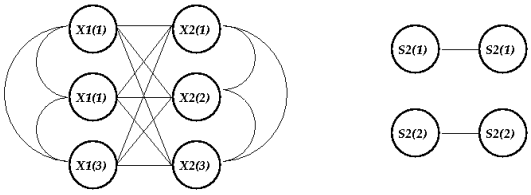


Diagram A: Graphical Models: (a) fully-connected graph denoting no independence assumptions (b) the ICA-factorized model with pair-wise only dependencies

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.

Diagram A 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 linkes). 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]

2.3 Distance-Sensitive Histograms

For the joint distribution estimation of k-tuples, we propose that not all the tuples should contribute equally to the histograms. We argue that an object's local appearance or structure is best captured by distance-sensitive histograming, in which the increment contributed by each tuple into a histogram bin depends upon the spatial adjacency structure among them.

For objects with fine-grain texture or structure, a larger increment should be added to the histogram for tuples with mutual distances on the order of the pattern periodicity. Conversely, for objects with distinct outer boundary structure, tuples with distances comparable to the object size are most representative of appearance and these should be given higher weights.

[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.

For k = 2, denoting the distance of the pair as d, the alternative methods are *inverse-distance-weighted (IDW) histograming*,

$$\Delta = e^{-\frac{d^2}{\sigma}} \tag{6}$$

or distance-weighted (DW) histograming,

$$\Delta = 1 - e^{-\frac{d^2}{\sigma}} \tag{7}$$

or simple hard-thresholding.

$$\Delta = \begin{cases} 1, & if \quad d \geq threshold \\ 0, & if \quad d < threshold \end{cases} \tag{8}$$

for differently structured images/objects.

3. IMPLEMENTATION ISSUES

To deal with noise as well as small variations in pose and lighting, the model histogram is passed through a Gaussian smoothing filter of variable sizes to achieve different trade-offs between accuracy and robustness.

In image database applications, the meta-data are usually extracted beforehand. To make the histograms from different images comparable, consistent quantization boundaries (bin width, bin range, etc) should be used across images. For some features such as color this is not an issue; while for others with large dynamic range, such as filter-bank responses, one must exercise extra caution to maintain histogram resolution and coverage. We used a large collection of images to estimate the range and frequency and cut 3-5% of the distribution tails before the quantization. This can improve the resolution of the histograms by over 100% in some cases with relatively little information loss.

4. 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.

4.1 Object detection/localization

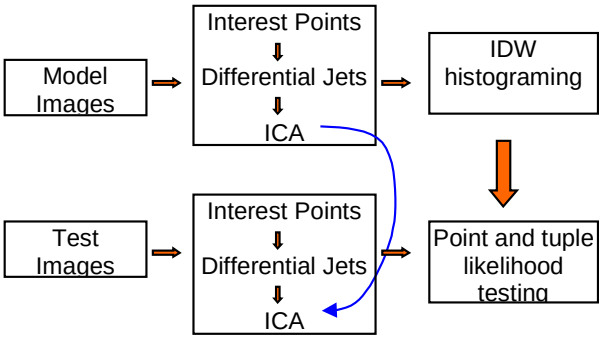
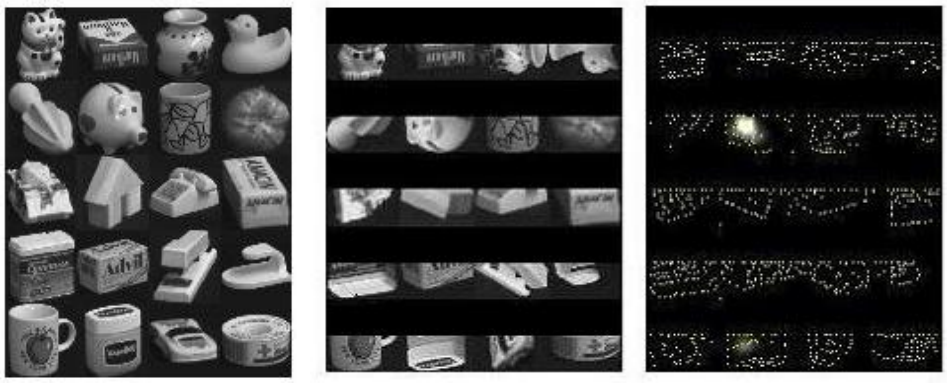


Figure 2. Diagram for object detection and localization (arrow indicates the same ICA basis is used)



(a) (b) (c)

Figure 3 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.

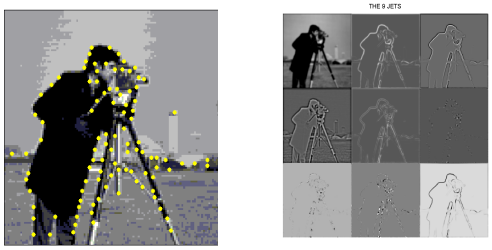


Figure 4 Harris interest point detections (left) and the 9 differential invariant Gaussian jets at all pixels (right)

First, tests on object detection in cluttered scenes were conducted. Figure 2 shows the flow diagram for this task. In Figure 2, 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.

4.1.1 Local feature extraction

For our experiments, we used a Harris operator [6][14] to detect interest points and extracted the first 9 differential invariant jets [9] at each point as the corresponding feature vector *x*. An example is shown in

Figure 4. 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. 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. *Inverse-distance-weighted* (IDW) histograming was applied in our experiments.

Test images were constructed using 20 objects from the Columbia Object Image Library (COIL) [12] (Figure 3(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 3(b)). Figure 3(c) shows the raw output for “piggy bank” detection on (b) where high-likelihood points have higher intensity.

4.1.2 Evaluation of Factorization

The effectiveness of ICA was evaluated by comparing 1 through 9 IC’s with the original 9 jets used as the feature vector. Note that for the original 9 jets, the histogram factorization along feature components is no longer valid, since the independence assumption on the differential invariant 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 improvement obtained by using the first 3 ICs in detecting the car (the original non-factorized 9-Djets confuse this object with the vaseline bottle).

By using 3 IC’s the system achieved 100% “first guess” detection (average rank = 1) on Figure 3(a), and an averaged rank of 1.2 for Figure 3(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. However since the features we used are not inherently invariant to *scale changes*, for a scale change of 50% the average rank reached 6.45. One possible solution for achieving scale invariance is to build object models at multiple scales.

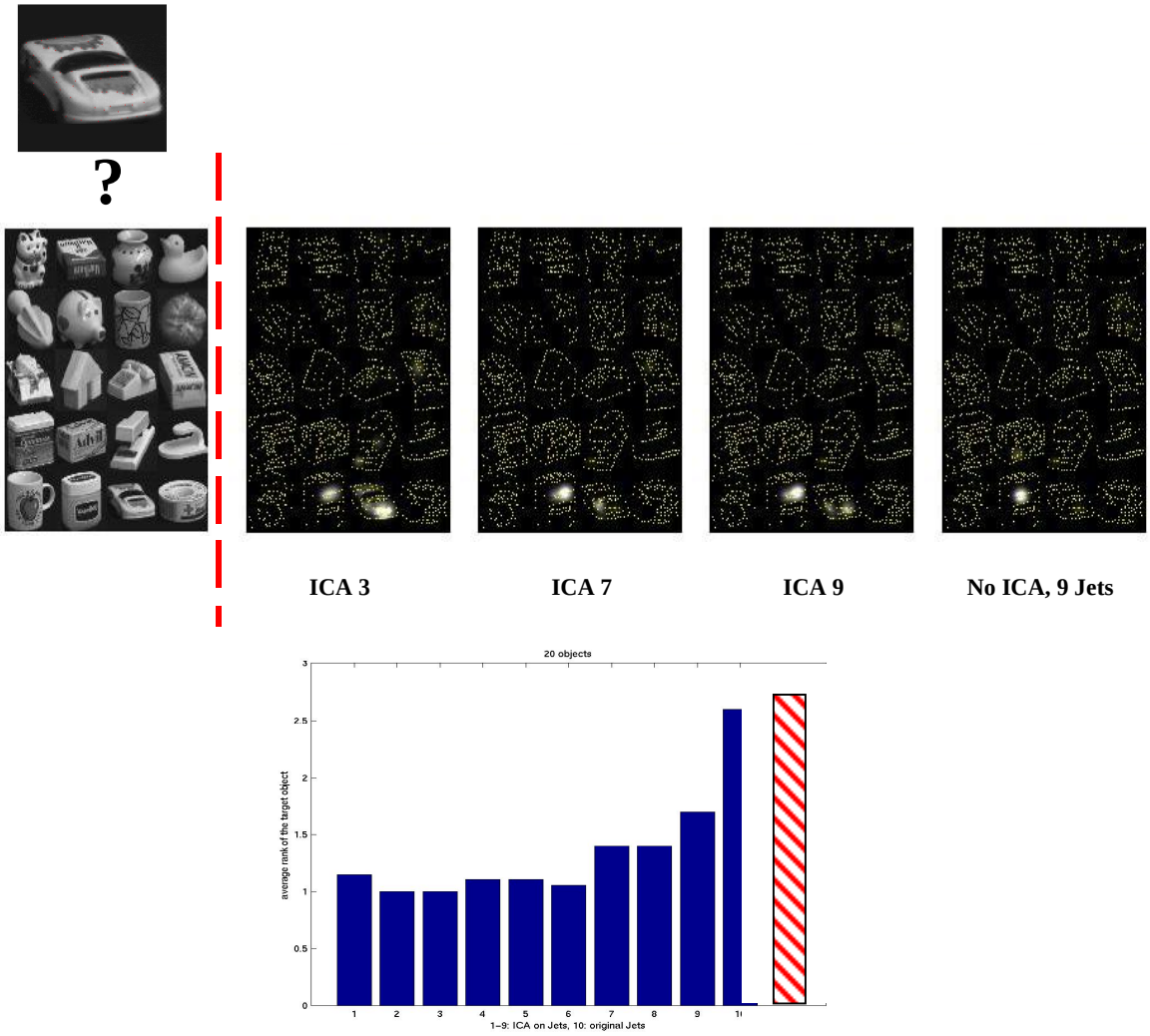


Figure 5 The upper part shows an example of detecting the car using ICA of 3, 7, 9 independent components as compared with no ICA; The lower part is the average detection rank of the target object using ICs (m = 1, 2, … 9) vs. original 9-dimensional jets (shown as the rightmost bar). Dataset: COIL; 20 objects

4.1.3 Modeling Object Classes

It is necessary to test object detection performance with greater variations such as that presented in Figure 6. Here we tested the detection of "leopard" and "tiger" on three images. Since we used window sizes of about 10 pixels for selecting interest points and jet computation, which is small compared to the image size and object sizes, this test is essentially equivalent to putting these images together as one cluttered scene.

In Figure 6, first a single model image of a leopard was used. The likelihood map, normalized to the range [0,1], was multiplied by the test images to highlight the high-probability regions. Shown in part (c) are the detection results for leopard: the detection maps reveal a high likelihood region in the first test image. It is also possible to form object class models by simply combining histograms from several training images. For example, we used six tiger images as training data and simply averaged their histograms to obtain a model for the class "tiger", which proved as effective as a single "prototype" model. In part (e), several high likelihood regions are detected in the third test image around the face and the neck of the tiger.

4.2 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

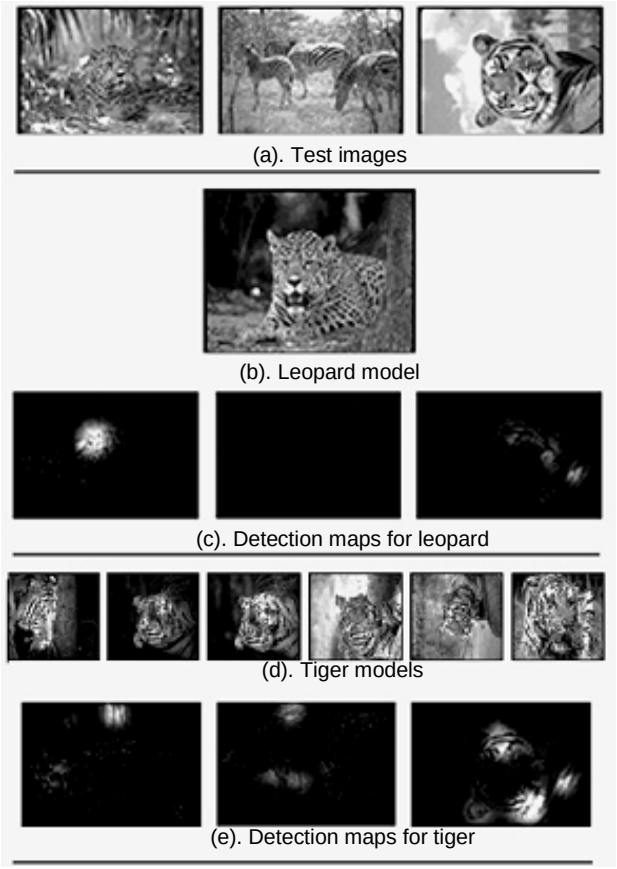


Figure 6 **Detecting Leopard and Tigers:** The likelihood maps are multiplied by the corresponding original images to reveal the detected (high likelihood) local structure.

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 Comparing our local/spatial apperance model with global features in image retrieval

Data set	COIL	COREL
Global Texture/Structure	96%	91%
Multi-Jet + Spatial HW	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 these two data sets.

5 DISCUSSION

We proposed a novel probabilistic modeling scheme based on factorization of high-dimensional distributions of local image features. Distance-sensitive *k*-tuple histograming was used for capturing local spatial dependencies. Our model exhibits an advantageous flexibility in modeling spatial relationships and can mediate a trade-off between non-rigid object modeling and distribution accuracy. Experiments have yielded promising results in image retrieval as well as in robust object localization in cluttered scenes.

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 Patter 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