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

ICA-based Probabilistic Local Appearance Models

Xiang Zhou, Baback Moghaddam, Thomas Huang

TR2001-29 December 2001

Abstract

This paper proposes a novel image modeling scheme for object detection and localization. Object appearance is modeled by the joint distribution of k-tuple salient point feature vectors which are factorized component-wise after an independent component analysis (ICA). Also, we propose a distance-sensitive histograming technique for capturing spatial dependencies. The advantages over existing techniques include the ability to model non-rigid objects (at the expense of modeling accuracy) and the flexibility in modeling spatial relationships. Experiments show that ICA does improve modeling accuracy and detection performance. Experiments in object detection in cluttered scenes have demonstrated promising results.

International Conferent on Image Processing (ICIṔ01), Thessaloniki, Greece, October 8-10th, 2001

Released July 2001.

ICA-based Probabilistic Local Appearance Models

Xiang Sean Zhou[1], Baback Moghaddam[2], Thomas S. Huang[1]
[1]*Beckman Institute for Advanced Science and Technology*
University of Illinois at Urbana Champaign, Urbana, IL 61801, USA
[2]*Mitsubishi Electric Research Laboratory, Cambridge, MA 02139, USA*
xzhou2@ifp.uiuc.edu, baback@merl.com, huang@ifp.uiuc.edu

Abstract

This paper proposes a novel image modeling scheme for object detection and localization. Object appearance is modeled by the joint distribution of k-tuple salient point feature vectors which are factorized component-wise after an independent component analysis (ICA). Also, we propose a distance-sensitive histograming technique for capturing spatial dependencies. The advantages over existing techniques include the ability to model non-rigid objects (at the expense of modeling accuracy) and the flexibility in modeling spatial relationships. Experiments show that ICA does improve modeling accuracy and detection performance. Experiments in object detection 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[14], or for local features and their spatial relationships [10][1][13][5]. 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 [13][10][9]; while others employ stochastic models and use probabilistic distance/matching metrics [5][8][1].

In this paper 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 Schneiderman and Kanade [13], 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 independent component analysis to reduce the dimensionality dramatically, 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 Huang et al. [5] or Chang and Krumm [1], the distance information is explicitly captured into the histogram bins. We argue in favor of collapsing the distance axis and instead using distance–dependent weights on the histogram increments. 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 can be a better choice. Again, this should be an application-dependent choice.

In this paper we will focus our attention only on the modeling of images/objects through 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 can be applied toward image retrieval or object detection in cluttered scenes. In Section 4 we present some preliminary results with discussions.

2. The Proposed Modeling Scheme

We propose joint multi-dimensional histograms as a non-parametric approximation of the joint distribution of image features at multiple image locations.

2.1 Classification by Class-conditional Density

Let i be the index for elementary feature components in an image, which can be pixels, corner/interest points [3][4], 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 [7][10][11] or even transform domain

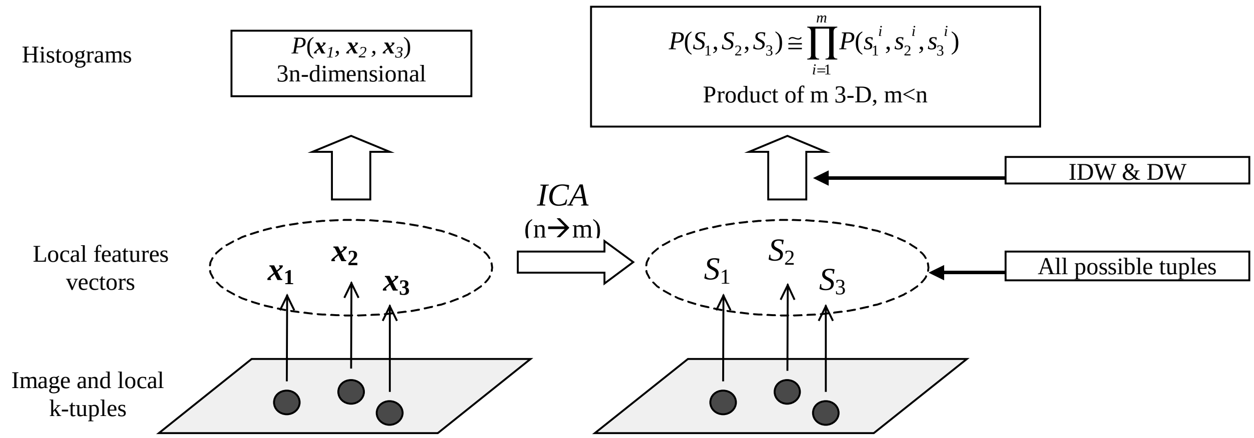


Figure 1 Image local appearance modeling by joint histograms

coefficients at an image block, or any other local/ regional features.

It is well known that maximum *a posteriori* decision rule, i.e.,

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

where M_l is the model and $T = \{\boldsymbol{x}_i\}$ is a test image, gives the minimum classification error. By Bayes theorem, and assuming equal priors, this is equivalent to maximum likelihood testing:

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

For the class-conditional density in (2), it is intractable to model dependencies among all $\boldsymbol{x}_i$'s, yet to completely ignore them will severely limit our modeling power—more often than not, objects distinguish themselves not through single patches of features, but by the interaction of different patches in a specific way. Without making a binary choice, we strike a balance between the two extremes by using joint histograms for *k*-tuples.

2.2 Joint distribution for *k*-tuples

In our system, we are not actually modeling the total joint likelihood of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_I$, which is an $(I\times n)$-dimensional distribution. Instead, we model the distribution of all *k*-tuples as an approximation:

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

Now this becomes a $(k\times n)$-dimensional distribution, which is still unworkable—e.g., for 20 histogram bins along each dimension, we have $20^{(k\times n)}$ bins to fill in. Therefore, we need to 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 independent. This is where independent component analysis (ICA) comes in.

2.3 Histogram factorization based on ICA

ICA originated in the context of blind source separation[6][2] to separate "independent causes" of a complex signal. 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

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

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. Without ICA the model suffers since in general the 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 results combined to form an overall score.

2.4 Distance-Sensitive Histograming for Modeling Spatial Dependencies

For the joint distribution estimation of *k*-tuples, not all the tuples are counted equally. 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 its 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. However, 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.

For the case of 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.

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

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 differential invariant Gaussian jets, one must exercise extra caution to maintain histogram resolution. We used a large collection of images to estimate the range and frequency and cut 3-5% of the tails before the quantization. This can improve the resolution of the histograms by over 100% in some cases with relatively little information loss.

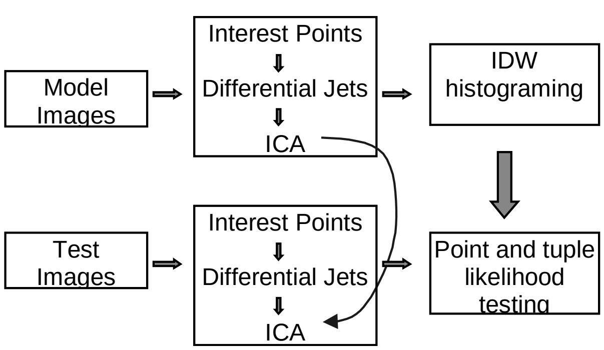
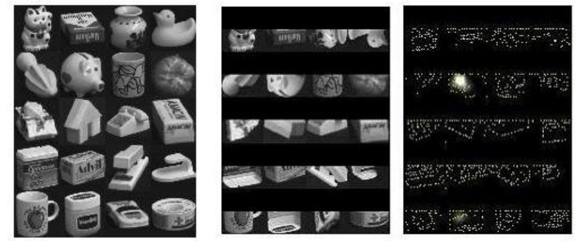


Figure 2. Diagram for the object detection and localization task implemented in this paper



(a) (b) (c)

Figure 3. Synthetic test images 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.

4. Experiments and Discussions

For our experiments, we used a Harris operator [4][11] to detect interest points and extracted the first 9 differential invariant jets [7] at each point as the corresponding feature vector $\boldsymbol{x}$. 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.

Tests on object detection in cluttered scenes were conducted in our study.

Figure 2 shows the flow diagram for this task. 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.

Test images were constructed using 20 objects from the Columbia Object Image Library (COIL) (Figure 3a). To test the invariance properties, each of the objects is transformed by pose change, a planar rotation, followed by 50% occlusion (Figure 3b). Figure 3c shows the raw output for "piggy bank" detection on b.

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

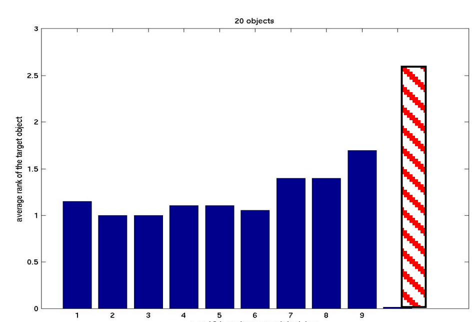


Figure 4. The average detection rank of the target object using ICs (m = 1, 2, … 9) vs. original 9-dimensional jets (shown as the rightmost bar). Dataset: COIL; 20 objects

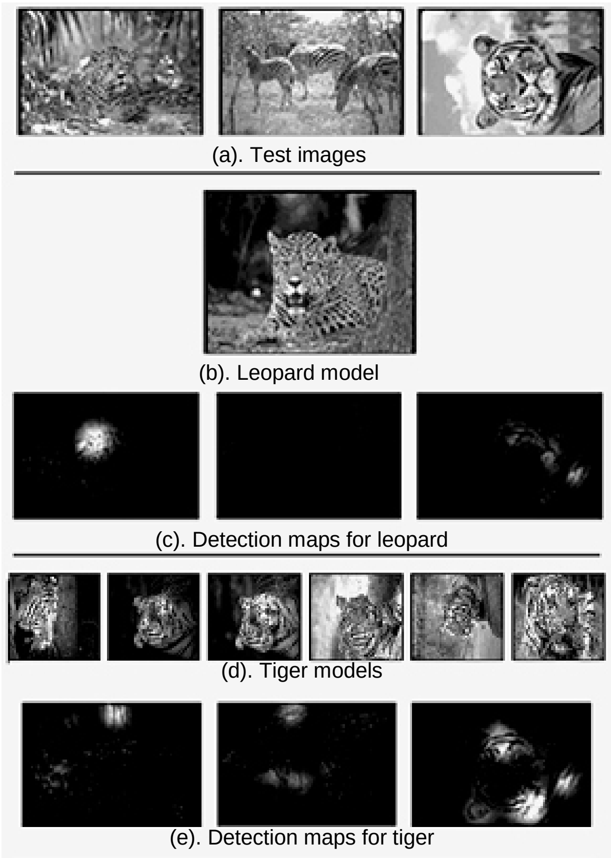


Figure 5 Detecting Leopard and Tigers
The likelihood maps are multiplied by the corresponding original images to reveal the detected (high likelihood) local structure.

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 4 depicts the clear improvement introduced by ICA. In fact, by only 3 IC's the system achieved 100% "first guess" detection (average rank = 1) on Figure 3a, and an averaged rank of 1.2 for Figure 3b.

It is necessary to test object detection performance with greater variations such as that presented Figure 5. 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 5, first a single model image of a leopard was used. The likelihood map, normalized to the range [0,1], was multiplied by the original 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. Second, we used six tigers as training images and simply averaged their histograms to obtain a model for "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.

5. Conclusion

A novel probabilistic image modeling scheme was proposed based on factorization of high-dimensional distributions of image features. We argued in favor of the distance-sensitive *k*-tuple histograming scheme for the purpose of capturing local spatial dependencies. In contrast to existing methods, the new scheme tries to mediate a trade-off between the capability for non-rigid object modeling and modeling accuracy. Another advantage of the proposed method is the flexibility in modeling spatial relationships. Experiments yield promising results on robust object localization in cluttered scenes.

References

[1] P. Chang, J. Krumm, "Object recognition with color cooccurrence histograms", Proc. IEEE Computer Vision and Pattern Recognition (CVPR'99), Colorado, Jun. 1999

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

[3] 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

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

[5] J. Huang, S. R. Kumar, M. Mitra, W.-J., Zhu, and R. Zabih, "Image indexing using color correlograms." Computer Vision and Pattern Recognition, (CVPR'97)

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

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

[8] 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.

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

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

[11] C. Schmid, R. Mohr, and C. Bauckhage, "Comparing and evaluating interest points", Proc. ICCV, 1998

[12] C. Schmid, "A structured probabilistic model for recognition", Computer Vision and Pattern Recognition (CVPR'1999), Fort Collins, Colorado, Jun. 1999

[13] H. Schneiderman, T. Kanade "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition." Computer Vision and Pattern Recognition (CVPR'98), pp. 45-51. 1998. Santa Barbara, CA.

[14] M. J. Swain and D. H. Ballard, "Color Indexing," Int'l Journal of Computer Vision, vol. 7, pp. 11-32,1991