HIERARCHICAL IMAGE PROBABILITY (HIP) MODELS

Clay Spence, Lucas Parra and Paul Sajda

Sarnoff Corporation
CN5300
Princeton, NJ 08543-5300

ABSTRACT

We formulate a model for probability distributions on image spaces. We show that any distribution of images can be factored exactly into conditional distributions of feature vectors at one resolution (pyramid level) conditioned on the image information at lower resolutions. We would like to factor this over positions in the pyramid levels to make it tractable, but such factoring may miss long-range dependencies. To capture long-range dependencies, we introduce hidden class labels at each pixel in the pyramid. The result is a hierarchical mixture of conditional probabilities, similar to a hidden Markov model on a tree. The model parameters can be found with maximum likelihood estimation using the EM algorithm. We have obtained encouraging preliminary results on the problems of detecting various objects in SAR images and target recognition in optical aerial images.

1. INTRODUCTION

Many approaches to object recognition in images estimate $\Pr(C \mid I)$, the probability that an object of class C is present in an image I. By contrast, a model of the probability distribution of images, $\Pr(I \mid C)$, has many attractive features. We could use this for object recognition in the usual way by training a distribution for each object class and using Bayes' rule to get $\Pr(C \mid I)$, or by using the likelihood ratio between $\Pr(I \mid C)$ and $\Pr(I \mid \bar{C})$. Clearly there are many other uses for image distributions, since any kind of data analysis task can be approached using knowledge of the distribution of the data. For classification we could attempt to detect unusual examples and reject them, rather than trusting the classifier's output. We could also compress, segment, interpolate, suppress noise, extend resolution, fuse multiple images, etc.

Many image analysis algorithms use probability concepts, but few treat the distribution of images, e.g., maximum entropy modeling [1]. There are several approaches that do not model the probability distribution on an image space, but motivated our work, e.g., MRF models [2, 3], the flexible histogram approach [4, 5], and multiscale stochastic processes [6]. All of these methods seem to be well-suited for modeling texture, but it is unclear how we might use them to capture the appearance of more structured objects.

As in many other approaches, we model the distribution of local image structure by using some local features, namely the outputs of some filters, and capture longer-range (either in scale or position) dependencies by modeling the influence of neighboring structures on each other. However, we argue that the presence of objects in images can make local conditioning like this inadequate. We capture these long-range dependencies by using hidden variables. The dependencies between the hidden variables in our model are local, like those in some MRF models, but marginalizing over them introduces long-range dependencies. We expect that such hidden variables would give poor models of object structure if they were only implemented at one pyramid level. Therefore we introduce them at all levels in a pyramid, and give them coarse to fine dependence.

2. THE HIP MODEL

To show that such a model can be a proper distribution on an image space, we show that any distribution on an image space can be factored into a coarse to fine hierarchy of conditional distributions. From an image I we build a Gaussian pyramid. Call the l-th level I_l, e.g., the original image is I_0. From each Gaussian level I_l we extract some set of feature images $\mathbf{F}_l$ (Figure 1). Sub-sample these to get feature images $\mathbf{G}_l$, so that the images in $\mathbf{G}_l$ have the same dimensions as I_{l+1}. Denote the set of images $\{I_{l+1}, \mathbf{G}_l\}$ by $\tilde{\mathbf{G}}_l$, and the mapping from I_l to $\tilde{\mathbf{G}}_l$ by $\tilde{\mathcal{G}}_l$. If $\tilde{\mathcal{G}}_l$ is invertible for all $l \in \{0, \ldots, L-1\}$ it is easy to show that

$$\Pr(I) = \left[\prod_{l=0}^{L-1} |\tilde{\mathcal{G}}_l| \Pr(\mathbf{G}_l \mid I_{l+1})\right] \Pr(I_L) \qquad (1)$$

In order to factor $\Pr(\mathbf{G}_l \mid I_{l+1})$ over positions, we introduce hidden variables. There is enormous freedom in this choice, although different choices can be easier or harder

We thank Jeremy De Bonet and John Fisher for kindly answering questions about their work and experiments. This work supported by the United States Government.

0-7803-6297-7/00/$10.00 © 2000 IEEE

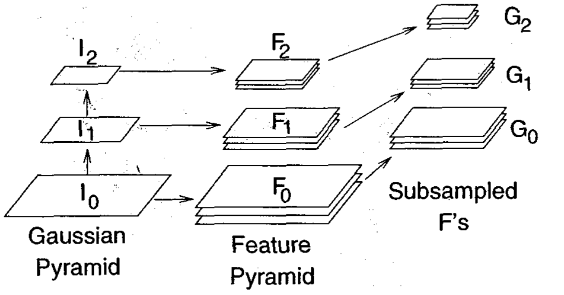

Fig. 1. Pyramids and feature notation.

to work with. One simple but non-trivial choice is to introduce an image A_l of integers at each level l. We assume that these contain enough information to allow us to factor $\Pr(\mathbf{G}_l \,|\, I_{l+1})$. Furthermore we assume that the local hidden variable $a_l(x)$ and the local lower-resolution feature vector $\mathbf{f}_{l+1}(x)$ carry all of the information in I_{l+1} that is relevant to the local feature vector $\mathbf{g}_l(x)$. This gives

$$\Pr(I) \propto \sum_{A_0,\ldots,A_{L-1}} \prod_{l=0}^{L} \prod_{x \in I_{l+1}} \Big[\Pr(\mathbf{g}_l \,|\, \mathbf{f}_{l+1}, a_l, x) \times \Pr(a_l \,|\, a_{l+1}, x)\Big], \quad (2)$$

where $a_{l+1}(x)$ is the hidden variable at the *parent* of x in the tree structure given by the sub-sampling operation. (To avoid repeating the string "(x)", we specify the location x as a conditioning variable in each $\Pr()$.)

3. TRAINING WITH EM

This model can be fit to data using an Expectation-Maximization (EM) algorithm. The E-step is the sum over hidden variables, which is tractable thanks to the tree structure of their dependencies. We choose $\Pr(\mathbf{g}_l \,|\, \mathbf{f}_{l+1}, a_l)$ to be normal with a mean that depends linearly on $\mathbf{f}_{l+1}$, i.e., $\Pr(\mathbf{g}_l \,|\, \mathbf{f}_{l+1}, a_l) = \mathcal{N}(M_{a_l}\mathbf{f}_{l+1} + \bar{\mathbf{g}}_{a_l}, \Lambda_{a_l})$. This makes the M-step tractable, and is rich enough to reproduce the non-Gaussian distribution of neighboring features on each other (see [7]). To enforce normalization we parameterize the label probabilities as $\Pr(a_l \,|\, a_{l+1}) = \pi_{a_l,a_{l+1}} / \sum_{a_l} \pi_{a_l,a_{l+1}}$. We denote by $\theta = \{\bar{\mathbf{g}}_{a_l}, M_{a_l}, \Lambda_{a_l}, \pi_{a_l,a_{l+1}}, \forall a_l, \forall l\}$ the vector of all parameters. For brevity we simply reproduce the relevant formulas without derivations.

To compute the expectations in the EM algorithm we need the joint probabilities of the image and individual labels at a position and pyramid level. These are given as

$$\Pr(a_l, a_{l+1}, x, I \,|\, \theta^t) = u_l(a_l, x)\tilde{d}_l(a_{l+1}, x)\Pr(a_l|a_{l+1}) \quad (3)$$

$$\Pr(a_l, x, I \,|\, \theta^t) = u_l(a_l, x) d_l(a_l, x), \quad (4)$$

where θ^t is the parameter vector from the t-th EM iteration. The quantities u and d are obtained through the upward and downward recursion relations

$$u_l(a_l, x) = \Pr(\mathbf{g}_l \,|\, \mathbf{f}_{l+1}, a_l, x) \prod_{x' \in \mathsf{Ch}(x)} \tilde{u}_{l-1}(a_l, x') \quad (5)$$

$$\tilde{u}_l(a_{l+1}, x) = \sum_{a_l} \Pr(a_l | \, a_{l+1}) u_l(a_l, x) \quad (6)$$

$$d_l(a_l, x) = \sum_{a_{l+1}} \Pr(a_l | \, a_{l+1}) \tilde{d}_l(a_{l+1}, x) \quad (7)$$

$$\tilde{d}_l(a_{l+1}, x) = \frac{u_{l+1}(a_{l+1}, \mathsf{Par}(x))}{\tilde{u}_l(a_{l+1}, x)} \, d_{l+1}(a_{l+1}, \mathsf{Par}(x)). \quad (8)$$

Here $\mathsf{Ch}(x)$ is the set of pixel locations in some level l that are children of pixel x in level $l+1$ in a tree relationship of pixels in the pyramid. Similarly, $\mathsf{Par}(x)$ is the parent pixel of x.

The upward recursion relations (5 – 6) is initialized at $l = 0$ with $u_0(a_0, x) = \Pr(\mathbf{g} \,|\, \mathbf{f}_1, a_0, x)$ and ends at $l = L$. At layer L (6) reduces to $\tilde{u}_L(a_{L+1}, x) = \tilde{u}_L(x)$.[1] Since we do not model any further dependencies beyond layer L, the pixels at layer L are assumed independent. The product of all $\tilde{u}_L(x)$ coincides with the total image probability, $\Pr(I|\theta^t) = \prod_{x \in I_L} \tilde{u}_L(x) = u_{L+1}$. The downward recursion (7 – 8) can be executed, starting with equation (8) at $l = L$ with $d_{L+1}(a_{L+1}, x) = d_{L+1}(x) = 1$.[1]

For the update equations, let us denote the average over position at level l weighted by $\Pr(a_l, x \,|\, I, \theta^t)$ by $\langle . \rangle_{t,a_l}$, i.e.,

$$\langle X \rangle_{t,a_l} = \frac{\sum_x \Pr(a_l, x \,|\, I, \theta^t) X(x)}{\sum_x \Pr(a_l, x \,|\, I, \theta^t)}. \quad (9)$$

Then the update equations for the Gaussian parameters are

$$M_{a_l}^{t+1} = \left(\left\langle \mathbf{g}_l \mathbf{f}_{l+1}^T \right\rangle_{t,a_l} - \langle \mathbf{g}_l \rangle_{t,a_l} \left\langle \mathbf{f}_{l+1}^T \right\rangle_{t,a_l}\right) \times \left(\left\langle \mathbf{f}_{l+1} \mathbf{f}_{l+1}^T \right\rangle_{t,a_l} - \langle \mathbf{f}_{l+1} \rangle_{t,a_l} \left\langle \mathbf{f}_{l+1}^T \right\rangle_{t,a_l}\right)^{-1}, \quad (10)$$

$$\bar{\mathbf{g}}_{a_l}^{t+1} = \langle \mathbf{g}_l \rangle_{t,a_l} - M_{a_l}^{t+1} \langle \mathbf{f}_{l+1} \rangle_{t,a_l}, \quad (11)$$

and

$$\Lambda_{a_l}^{t+1} = \left\langle \left(\mathbf{g}_l - M_{a_l}^{t+1} \mathbf{f}_{l+1}\right) \left(\mathbf{g}_l - M_{a_l}^{t+1} \mathbf{f}_{l+1}\right)^T \right\rangle_{t,a_l} - \bar{\mathbf{g}}_{a_l}^{t+1} \bar{\mathbf{g}}_{a_l}^{t+1\,T}. \quad (12)$$

The update equation for the label probability parameters is

$$\pi_{a_l,a_{l+1}}^{t+1} = \sum_x \Pr(a_l, a_{l+1}, x | I, \theta^t). \quad (13)$$

[1] The (non-existent) label a_{L+1} can be thought of as a label with a single possible value, which is always set. The conditional $\Pr(a_L \,|\, a_{L+1})$ turns then into a prior $\Pr(a_L)$

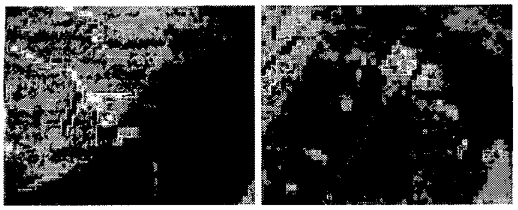

Fig. 2. Examples of positive (left) and negative (right) ROIs for the aircraft detection problem. Data from the MassGIS at `http://ortho.mit.edu/nsdi/`.

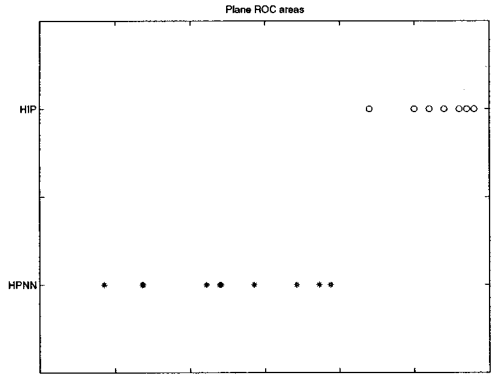

Fig. 3. A_z values from a jack-knife study of detection performance of HIP and HPNN (hybrid pyramid/neural network) models.

4. EXPERIMENTS

We have applied this HIP model to two problems. The first was to detect aircraft in aerial photographs. The HIP model performed substantially better than our own hybrid pyramid neural network (HPNN) algorithm [8]. (See Figures 2 and 3.) (For a better comparison we would select features independently for the HIP and HPNN models. The HPNN gave $A_z = 0.86$ with a different set of features.)

For vehicle discrimination in SAR, we performed an experiment with the three target classes in the MSTAR public targets data set, to compare with the results of the flexible histogram approach of De Bonet, et al [5]. We trained three HIP models, one for each of the target vehicles BMP-2, BTR-70 and T-72 (Figure 4). As in [5] we trained each model on ten images of its class, one image for each of ten aspect angles, spaced approximately 36° apart. We trained one model for all ten images of a target, whereas De Bonet et al trained one model per image.

We first discriminated between vehicles of one class and other objects by thresholding $\log \Pr(I \mid C)$, i.e., no model of other objects is used. For the tests, the other objects were taken from the test data for the two other vehicle classes, plus seven other vehicle classes. There were 1,838 image from these seven other classes, 391 BMP2 test images, 196 BTR70 test images, and 386 T72 test images. The resulting ROC curves are shown in Figure 5a.

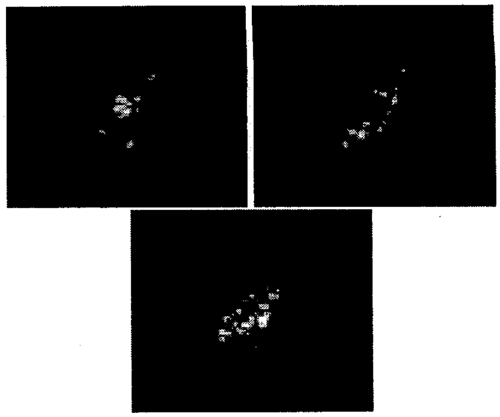

Fig. 4. SAR images of three vehicle classes. Data from the MSTAR public data set.

A second discrimination criterion that uses a distribution is the likelihood ratio, $\log \Pr(I \mid C_1) - \log \Pr(I \mid C_2)$. Here we cannot use the extra seven vehicle classes. The resulting ROC curves are shown in Figure 5b. The performance is comparable to that of the flexible histogram approach of De Bonet et al.

5. CONCLUSION

We have presented a hierarchical image probability (HIP) model for probability distributions of images, and demonstrated its utility in a pair of object recognition tasks. The model uses hidden class labels to capture long-range dependencies. A distribution model has many potential uses besides recognition, including compression, noise suppression, novelty detection, segmentation, etc.

The HIP model has two key elements. First is the restriction that the features be invertible to make the model a proper probability distribution on the image space. It appears to be possible to relax these restrictions in some cases. Second is the use of hidden variables, since these are needed to express long-range dependencies in the model. Our current hidden variable structure was chosen for tractability, since we can explicitly marginalize the hidden variables in this structure. Generalizations like choosing a connectivity denser than a tree, or including continuous hidden variables could have benefits, but we would need approximations to evaluate the probabilities. There is much room for further work along these lines.

We are also working on sampling from HIP models, i.e., generating random images. This capability provides an in-

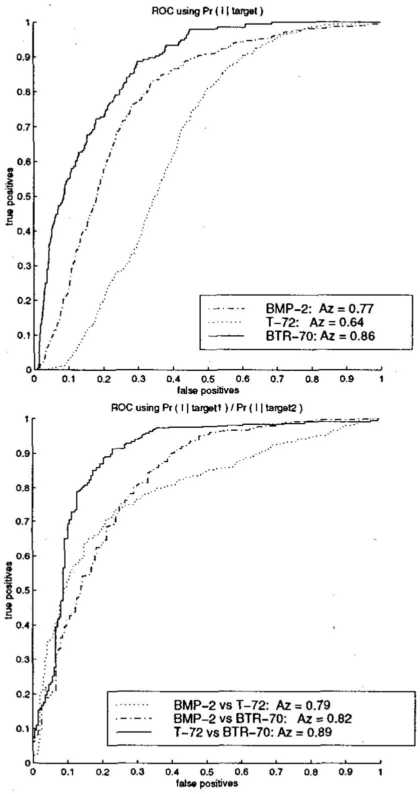

Fig. 5. ROC curves for vehicle detection in SAR imagery. (Upper: ROC curves by thresholding HIP likelihood of desired class. Lower: ROC curves for inter-class discrimination using ratios of likelihoods as given by HIP models. ,

dependent means of evaluating the model that is not available with neural network models of $\Pr(C \mid I)$.

6. REFERENCES

[1] Song Chun Zhu, Ying Nian Wu, and David Mumford, "Minimax entropy principle and its application to texture modeling," *Neural Computation*, vol. 9, no. 8, pp. 1627–1660, 1997.

[2] Stuart Geman and Donald Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. PAMI*, vol. PAMI-6, no. 6, pp. 194–207, Nov. 1984.

[3] Rama Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. ASSP*, vol. 33, pp. 959–963, 1985.

[4] Jeremy S. De Bonet and Paul Viola, "Texture recognition using a non-parametric multi-scale statistical model," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 1998.

[5] J. S. De Bonet, P. Viola, and J. W. Fisher III, "Flexible histograms: A multiresolution target discrimination model," in *Proceedings of SPIE*, E. G. Zelnio, Ed., 1998, vol. 3370.

[6] Mark R. Luettgen and Alan S. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination," *IEEE Trans. Image Proc.*, vol. 4, no. 2, pp. 194–207, 1995.

[7] Robert W. Buccigrossi and Eero P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," Tech. Rep. 414, U. Penn. GRASP Laboratory, 1998, Available at ftp://ftp.cis.upenn.edu/pub/eero/buccigrossi97.ps.gz.

[8] Clay D. Spencė and Paul Sajda, "Applications of multiresolution neural networks to mammography," in *Advances in Neural Information Processing Systems 11*, Michael S. Kearns, Sara A. Solla, and David A. Cohn, Eds., Cambridge, MA, 1998, pp. 981–988, MIT Press.