

Using texture to analyze and manage large collections of remote sensed image and video data

Shawn Newsam, Lei Wang, Sitaram Bhagavathy, and Bangalore S. Manjunath

We describe recent research into using the visual primitive of texture to analyze and manage large collections of remote sensed image and video data. Texture is regarded as the spatial dependence of pixel intensity. It is characterized by the amount of dependence at different scales and orientations, as measured with frequency-selective filters. A homogeneous texture descriptor based on the filter outputs is shown to enable (1) content-based image retrieval in large collections of satellite imagery, (2) semantic labeling and layout retrieval in an aerial video management system, and (3) statistical object modeling in geographic digital libraries. © 2004 Optical Society of America

OCIS codes: 100.2960, 100.5010, 110.2960.

1. Introduction

Remote sensed data, such as satellite imagery or aerial videography, are being acquired at increasing rates because of technological advances in airborne and spaceborne optical sensing systems. The full value of these data sets is not being realized, however, because of the prohibitive cost of manual analysis, both in terms of time and money. It has been hoped that automated analysis methods with computers would help clear this information bottleneck. Although progress has been made, the rate of analysis still lags the rate of acquisition. Further advancements in the automated analysis of remote sensed data sets are urgently needed.

Most research into automating the analysis has focused on the spectral dimension. This has been the case even though the data are intrinsically spatial. Other research has primarily focused on exploiting the spatial dimension to localize the analysis or detect well-structured objects.^{1,2} The research presented in this paper represents our recent efforts toward using spatial information to help automate the analysis and management of remote sensed image and video data.

The basis of the approach is to consider the spatial

relationships of pixel intensities in remote sensed image and video data as the visual primitive of texture. The analysis of image texture has received significant research attention during the past several decades, perhaps only second to color.³⁻⁸ Yet developing effective descriptors for such a fundamental image feature remains a challenge.

There are generally two approaches to analyzing texture quantitatively. The first approach explicitly analyzes the spatial dependence of pixel intensity in the texture. Examples of this approach include co-occurrence matrices⁴ and methods based on Markov random fields (MRFs).⁵⁻⁷ In the second approach the structure in the texture is inferred by analyses of the distribution of frequencies with appropriate filters. The former is more precise and intuitive in the visual characterization of the texture. However, in practice, it fails to accommodate the pixel variations among textures that are considered similar. The latter is more useful for similarity retrieval because it is a global approximation of the image structure. It is the approach taken by the research presented in this paper. In particular, orientation and scale-selective filters are used to characterize what can be considered as the direction and coarseness of a texture.

We begin this paper by describing the Gabor filters that form the basis of the texture analysis. We provide details on how orientation and scale-selective filters are constructed by varying the modulation of Gabor functions. In the remainder of the paper we describe how the outputs of Gabor filter banks enable texture-based analysis and management of remote sensed imagery. Specifically, a homogeneous tex-

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, California 93106. The e-mail address for S. Newsam is snewsam@ece.ucsb.edu.

Received 20 May 2003; revised manuscript received 7 July 2003.
0003-6935/04/020210-08\$15.00/0

© 2004 Optical Society of America

ture descriptor recently standardized by the Moving Picture Experts Group (MPEG) is used to perform content-based similarity retrieval in large collections of aerial and satellite imagery. The descriptor also provides a semantic labeling of aerial videography in a video database management system. This labeling enables similarity retrieval based on the semantic layout of the video frames. Finally, the filter outputs enable an object-based representation for remote sensed imagery. In particular, the characteristic textures, or texture motifs, of objects that defy traditional modeling approaches are learned by use of a statistical approach.

2. Texture Analysis with Gabor Filters

Use of filters based on Gabor functions to analyze texture is motivated by several factors. First, the Gabor representation can be shown to be optimal in the sense that it minimizes the joint two-dimensional uncertainty in space and frequency.⁹ Second, the fact that Gabor functions can be used to model the receptive fields of simple cells in the mammalian visual cortex¹⁰ is psychovisual evidence that Gabor-like filtering takes place early on in the human visual system.

A bank of orientation and scale-selective filters are constructed as follows.⁸ A two-dimensional Gabor function $g(x, y)$ and its Fourier transform $G(u, v)$ can be written as

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W x \right], \quad (1)$$

$$G(u, v) = \exp \left\{ -\frac{1}{2} \left[\frac{(u - W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\}, \quad (2)$$

where $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$. A class of self-similar functions referred to as Gabor wavelets is now considered. Let $g(x, y)$ be the mother wavelet. A filter dictionary can be obtained by appropriate dilations and translations of $g(x, y)$ through the generating function:

$$\begin{aligned} g_{rs}(x, y) &= a^{-s}g(x', y'), \quad a > 1, \\ & \quad s \in 0, \dots, S - 1, r \in 1, \dots, R, \\ x' &= a^{-s}(x \cos \theta + y \sin \theta), \\ y' &= a^{-s}(-x \sin \theta + y \cos \theta), \end{aligned} \quad (3)$$

where $\theta = (r - 1)\pi/R$. The indices r and s indicate the orientation and scale of the filter, respectively. R is the total number of orientations and S is the total number of scales in the filter bank. The scale factor a^{-s} in Eqs. (3) is meant to ensure that the energy is independent of s .

Although the size of the filter bank is application dependent, experimentation has shown that a bank of filters tuned to combinations of five scales, at octave intervals, and six orientations, at 30-deg intervals, is sufficient for the analysis of remote sensed

imagery. See Ref. 8 for more details on the filter bank construction.

A. Homogeneous Texture Descriptor

The homogeneous texture descriptor is a compact representation of the Gabor filter outputs.¹¹ It is a feature vector formed when the first and second moments of the filter outputs are computed for a spatially homogeneous image or image region. Representing texture as a point in this multidimensional feature space is useful because closeness in the feature space turns out to correspond well with visual similarity. The visual similarity between images with respect to texture can be quantitatively computed with a distance function. These distances can be used to perform content-based similarity retrieval (examples to follow). Furthermore, machine learning techniques can be used to estimate the distributions of texture classes from manually labeled examples. We can label new images using these distributions.

The homogeneous texture descriptor is formed as follows. Suppose $f_{11}(x, y), \dots, f_{RS}(x, y)$ are the outputs of a filter bank tuned to R orientations and S scales. The feature vector f is then

$$f = [\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, \dots, \mu_{1S}, \sigma_{1S}, \dots, \mu_{RS}, \sigma_{RS}] \quad (4)$$

where μ_{rs} and σ_{rs} are the mean and standard deviation of $f_{rs}(x, y)$, respectively. This descriptor was recently standardized by the International Standards Organization MPEG-7 Multimedia Content Description Interface Standard¹² with the minor modification that the mean and standard deviation of the filter outputs are computed in the frequency domain for reasons of efficiency.

B. Similarity Retrieval

Low-level visual primitives have been used for some time to perform content-based similarity retrieval in image data sets.¹³⁻¹⁶ The homogeneous texture descriptor is particularly effective for similarity retrieval in remote sensed imagery. We compute the visual similarity between images by defining a distance measure in the $2RS$ -dimensional texture feature space. Euclidean distance is commonly used so that the similarity between images $I^{(1)}$ and $I^{(2)}$ is computed as

$$d[I^{(1)}, I^{(2)}] = \|f^{(1)} - f^{(2)}\|_2. \quad (5)$$

A query-by-example paradigm can be used to perform content-based similarity retrieval in image data sets. We compute the nearest-neighbor queries by returning those images with the least distance to the query. We compute the range queries by returning those images whose distances to the query image are smaller than some specified threshold.

To localize the descriptor, large satellite or aerial images are typically divided into tiles measuring 64×64 or 128×128 pixels. This tiling results in large data sets because of the size of the original



Fig. 1. Results of a range query in a satellite image. The right-most tile is the query tile. The other tiles are the results.

images. We compute the homogeneous texture descriptor for each tile by applying a Gabor filter bank tuned to combinations of five scales and six orientations. The output of the 30 filters is summarized by a 60-component feature vector, as described above. Similarity retrieval is performed when we compute the distances between a query tile and the rest of the data set.

Figure 1 shows an example of a range query for a data set of approximately 400,000, 64×64 pixel tiles from seven satellite images with 1-m resolution. The query tile is the rightmost tile on the freeway (the tiles are indicated with a white border). The other tiles are the results of a range query. Note that high precision and recall are achieved because only the remaining freeway tiles are retrieved. Figure 2 shows an example of a nearest-neighbor query for the same data set. The top-left tile is the query tile. The other tiles are those closest to the query among the 400,000 tile data set.

Euclidean distance results in a visual similarity measure that is orientation sensitive. Orientation-invariant similarity measurement is possible by use of the following distance function:

$$d_{RI}[I^{(1)}, I^{(2)}] = \min_{r \in R} \|f_{(r)}^{(1)} - f^{(2)}\|_2. \quad (6)$$

Here, $f_{(r)}$ indicates the feature vector f circularly shifted by r rotations. Conceptually, this distance function computes the best match between rotated versions of the images. Figure 3 shows an example of an orientation-invariant nearest-neighbor search

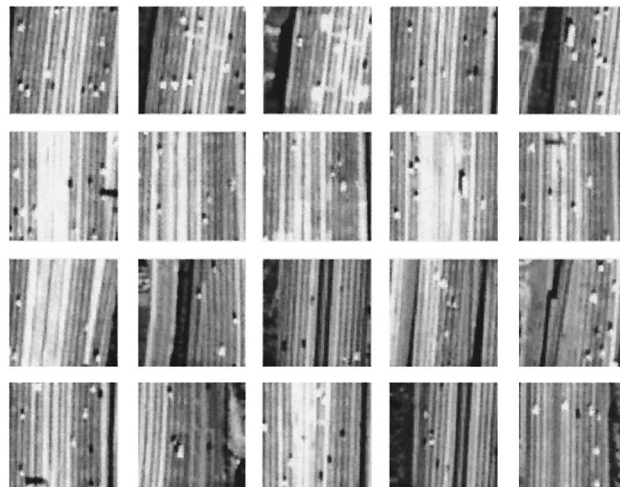


Fig. 2. Results of a nearest-neighbor query in a data set of satellite images. The top-left tile is the query tile. The other tiles are the results.

by use of the same query tile as in Fig. 2. Again, the top-left tile is the query tile.

Similarity retrieval examples such as these demonstrate that the descriptor provides meaningful annotation of remote sensed imagery. In fact, the descriptor was accepted to the MPEG-7 standard after extensive experimentation and testing with an evaluation data set that included a large collection of aerial images. A demonstration of content-based similarity retrieval by use of the MPEG-7-compliant homogeneous texture descriptor is available on line.¹⁷

3. Semantic Image Labeling

The above similarity retrieval examples demonstrate that the homogeneous texture descriptor provides an effective low-level perceptual description of remote sensed imagery. Low-level descriptions are limited, however, in their ability to support the kinds of higher-level interaction that are ultimately required.

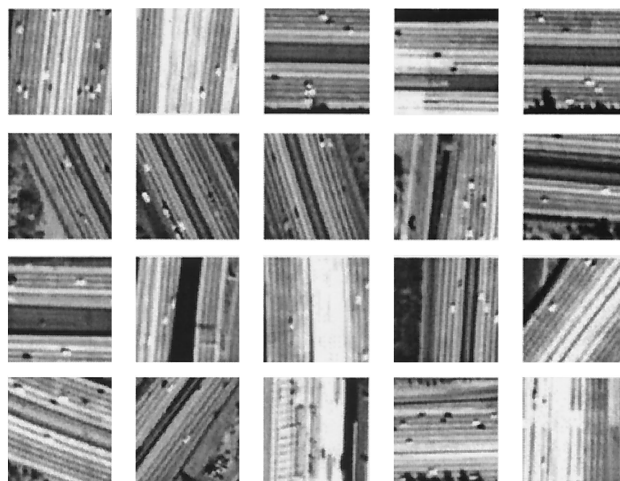


Fig. 3. Results of an orientation-invariant nearest-neighbor query. The top-left tile is the query tile. The other tiles are some of the results.

In this section we describe how statistical machine learning techniques can be used to extend this low-level description to provide a more meaningful semantic level of interaction.

The approach has two salient components. First, the statistical distributions of texture feature vectors corresponding to different semantic classes are learned. This allows novel image regions to be semantically labeled. Second, a MRF model is used to check for inconsistencies in the spatial arrangement of this labeling. The two components are implemented in a video management system that allows users to interact with aerial videography data sets based on the spatial layouts of video frames.

A. Semantic Labeling

The statistical distribution of texture features corresponding to different semantic classes, such as roads, buildings, and fields, can be learned with a labeled training set. In the proposed approach, this training set consists of image or video frame tiles that have been manually assigned semantic labels from a predetermined set of classes. The tiles are also annotated by use of the homogeneous texture descriptor. The premise of the approach is that, because (1) visually similar textures form clusters in the sparse feature space and (2) there is a wide variation in visual appearance within each semantic class, then Gaussian mixture models (GMMs) can be used to model the feature distributions conditioned on the class labels. In particular, under the assumption that each semantic class forms K clusters in the feature space, we model the distribution of features $Y \in R^d$ ($d = 60$) for a class as a mixture of K Gaussians using the following density function:

$$P(y|\theta_m) = \sum_{j=1}^K \alpha_{mj} \frac{1}{[(2\pi)^d |\Sigma_{mj}|]^{1/2}} \exp \left[-\frac{1}{2} (y - \mu_{mj})^T \Sigma_{mj}^{-1} (y - \mu_{mj}) \right], \quad (7)$$

where $\theta_m = \{\alpha_{mj}, \mu_{mj}, \Sigma_{mj}\}_{j=1}^K$ is the parameter set for the semantic class m . α_{mj} are the mixture coefficients, $\mu_{mj} \in R^d$ is the mean of the j th Gaussian, and Σ_{mj} is the covariance matrix of the j th Gaussian.

B. Spatial Layout Consistency

We labeled the image tiles using the set of trained GMMs and a maximum-likelihood classifier. This labeling is rarely perfect, however. An effective way to improve its accuracy is to impose constraints on the spatial arrangement of the labels. Inconsistencies can be ruled out, such as a tile labeled parking lot being surrounded by tiles labeled water. In this particular approach we use MRFs to model the spatial distribution of the class labels. The label of a block at site s is modeled as a discrete-valued random variable X_s , taking values from the semantic label set $\mathbf{M} = \{1, 2, \dots, M\}$, and the set of random variables $X = \{X_s, s \in S\}$ constitutes a random field where S is the

lattice of image blocks. The random field X is modeled as a MRF with a Gibbs distribution¹⁸:

$$p(x) = \frac{1}{Z} \exp[-U(x)], \quad (8)$$

where x is a realization of X . The Gibbs energy function $U(x)$ can be expressed as the sum of clique potential functions:

$$U(x) = \sum_{c \in Q} V_c(x), \quad (9)$$

where Q is the set of all cliques in a neighborhood. The MRF model is reinforced when we incorporate the class-conditioned feature likelihoods into the energy function, as follows:

$$U(x) = \sum_{s \in S} \left(\sum_{s' \in N_s} -\alpha LP_{s-s'} - \beta LP_s \right), \quad (10)$$

where $LP_{s-s'} = \log[p_{s-s'}(x_s, x_{s'})]$ and $LP_s = \log[p(y_s|\theta_{x_s})]$. N_s is the set of neighbors of the site s and α and β are the weights of $LP_{s-s'}$ and LP_s , respectively. $LP_{s-s'}$ represents the spatial relationship between neighboring sites s and s' where $s-s'$ indicates the direction of neighborhood. LP_s represents the conditional probability density of feature vector y_s given the label x_s . $p_{s-s'}(x_s, x_{s'})$ is the joint probability of x_s and $x_{s'}$ along the direction $s-s'$ and can be approximated with a co-occurrence matrix from the labeled training set. For each type of clique $s-s'$, a co-occurrence matrix is constructed from the joint probabilities $P_r(i, j)$ between pairs of semantic labels i and j in a given direction r .

To simplify the model, a second-order pair-site neighborhood system is used. Each site thus has eight neighbors. Four types of clique are considered, wherein $s-s'$ makes angles of 0, 45, 90, and 135 deg with respect to the x axis. In any neighborhood, cliques along the same direction are considered equivalent. Four co-occurrence matrices are constructed along these four directions of label distribution.

C. Spatial Layout Retrieval

In this step, a representation termed semantic layout is used to characterize an image or video frame based on the spatial arrangement of its labeled tiles. For retrieval purposes, the similarity between the query image and each stored image can be determined from their semantic layouts. Let the semantic layout of the query image be X^q and that of the stored image be X^l . To improve the retrieval performance, a soft classification scheme is adopted. For a given image tile, the labels with the three largest local conditional probabilities are selected to represent this tile. All candidate labels are stored along with the feature vectors for future retrieval. The modified semantic

layout similarity between the query image and each stored image is given by

$$\begin{aligned}
 S_3 = & \sum_{s \in S} \left[\sum_{j=1}^3 a_1 a_j \delta(x_{s,j}^q, x_{s,1}^I) \right] \\
 & + \sum_{s \in S} \left[\sum_{j=1}^3 a_2 a_j \delta(x_{s,j}^q, x_{s,2}^I) \right] \\
 & \times \left[1 - \sum_{i=1}^3 \delta(x_{s,i}^q, x_{s,1}^I) \right] \\
 & + \sum_{s \in S} \left[\sum_{j=1}^3 a_3 a_j \delta(x_{s,j}^q, x_{s,3}^I) \right] \\
 & \times \left[1 - \sum_{i=1}^3 \delta(x_{s,i}^q, x_{s,2}^I) \right] \\
 & \times \left[1 - \sum_{i=1}^3 \delta(x_{s,i}^q, x_{s,3}^I) \right], \quad (11)
 \end{aligned}$$

where $a_i = 1/2^{i-1}$, $i = 1, 2, 3$, are the weights for different label similarities, and $x_{s,j}^q$ is the j th candidate label of the query image tile at site s .

In Eq. (11), we computed the similarity measure by comparing each candidate label of a query image tile to all the candidate labels of the corresponding stored image tile. This approach to similarity retrieval is expected to perform better than existing methods that do not consider the underlying semantics.

The similarity measure in Eq. (11) is effective only when all images are of the same size. To compare the layouts of images of any size, a semantic histogram can be computed for each image. This is similar to the image histogram where the inputs are semantic labels instead of image intensities. As long as the semantic label set is the same, two different-sized images can be compared by use of their semantic histograms.

Examples of semantic layout retrieval in a prototype video management system are shown in Figs. 4 and 5. In each example, the top frame is the query and the remaining frames are the results. The semantic labeling of the query is shown. Observe that the retrieved frames can be visually quite different from the query even though their semantic layouts are similar. These results could not be achieved by a low-level description alone. The combination of GMM and MRF models is thus shown to capture the semantic layout of the frames.

4. Geospatial Object Modeling

So far, the texture descriptor has been used for similarity retrieval in large collections of remote sensed image or video data. In this section we describe how the texture features enable an object-based representation for remote sensed imagery.¹⁹ Such a representation allows users to interact with the data at a more intuitive level. In the proposed approach, geospatial objects that consist of multiple characteristic textures are modeled by mixtures of Gaussians. The model parameters are estimated with unsuper-

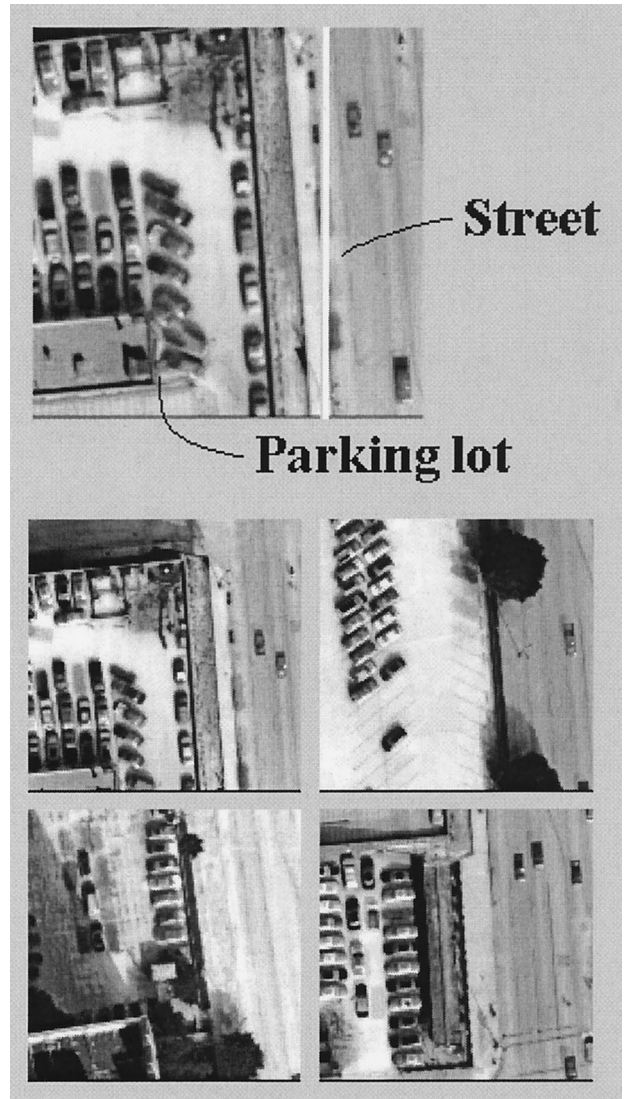


Fig. 4. Example of a spatial layout retrieval in a prototype video management system. The top frame is the query. The lower four frames are the results.

vised learning techniques (unlike in Section 3 where the parameters are estimated from labeled training sets). The goal is to use the models to locate and catalog new objects instances in remote sensed imagery.

The focus is on modeling geospatial objects that have characteristic textures but lack structure or well-defined boundaries. Examples of such objects include harbors, golf courses, and mobile home parks. These objects defy traditional object modeling approaches, such as template matching, so new approaches are needed. With the proposed technique we model the objects by statistically characterizing the common textures, or texture motifs, such as the rows of boats and water in harbors or the grass fairways and trees in golf courses.

The texture motifs are learned by use of the Gabor texture features extracted from sample object instances. An RS -dimensional feature vector c is ex-

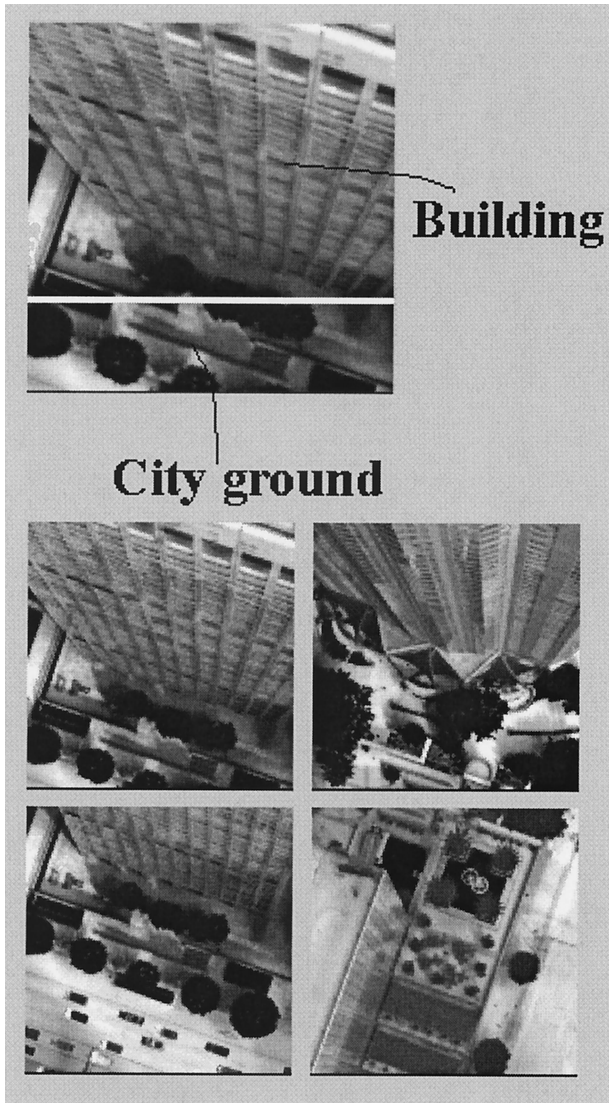


Fig. 5. Example of a spatial layout retrieval in a prototype video management system. The top frame is the query. The lower four frames are the results.

tracted for each pixel with a filter bank of R orientations and S scales. (Note that the analysis is done at the pixel level, not at the tile level.) Under the assumptions that (1) the pixels for a class of objects are generated by one of N possible texture motifs and (2) these motifs have Gaussian distributions in the feature space, the probability density function of c can be expressed as a mixture distribution:

$$p(c) = \sum_{j=1}^N P(j)p(c|j), \quad (12)$$

where $p(c|j)$ is the conditional likelihood of the feature c being generated by motif j , and $P(j)$ is the prior probability of motif j . The number of motifs N along with distribution means and covariance matrices are the parameters that completely specify the class model.

The expectation maximization algorithm²⁰ is used

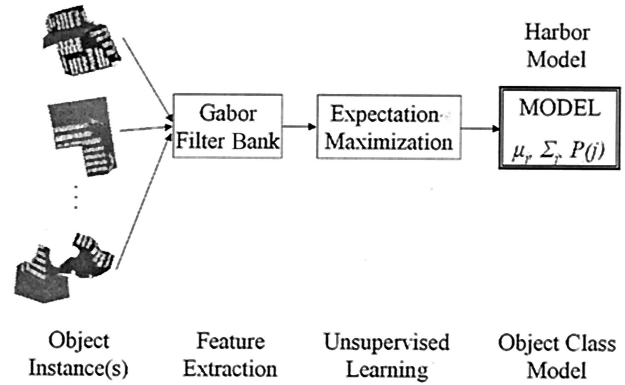


Fig. 6. Model parameters are learned from a sample set with the expectation maximization algorithm.

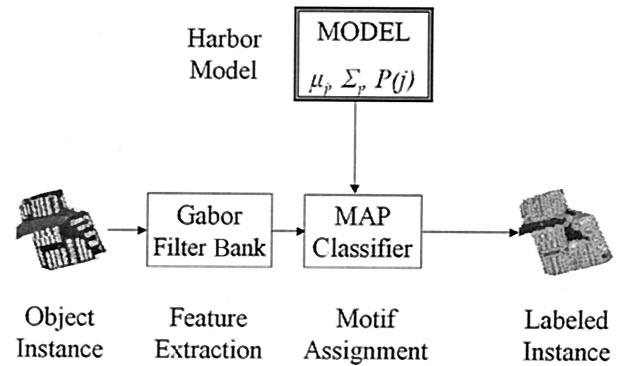


Fig. 7. Motif labeling with a MAP classifier.

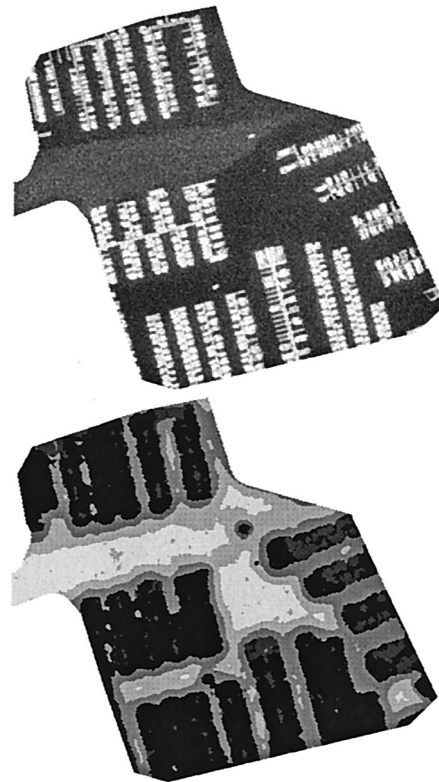


Fig. 8. Harbor image and its corresponding labeling. The gray level indicates the motif.

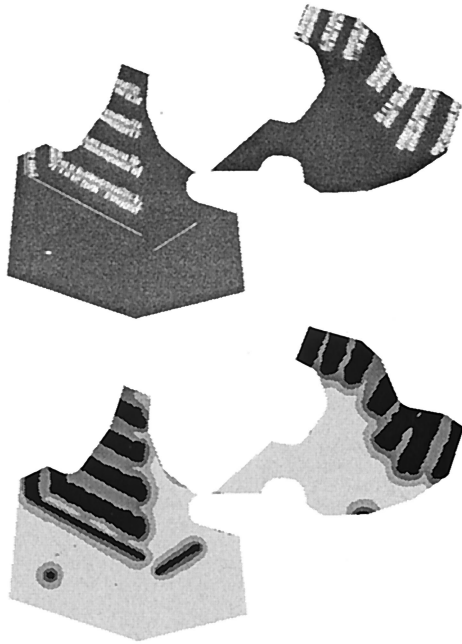


Fig. 9. Second harbor image and its corresponding labeling. The gray level indicates the motif.

to estimate the GMM parameters from a set of sample object instances, as shown in Fig. 6. It is important to note that the characteristic textures are determined solely from the distribution of the texture features.

A trained GMM can then be used to label the motifs for an object instance, as shown in Fig. 7. A maximum *a posteriori* (MAP) classifier assigns label i^* to a pixel with feature vector c according to

$$i^* = \arg \max_{1 \leq i \leq N} [P(i|c)]. \quad (13)$$

The posterior probabilities $P(i|c)$ are computed with the Bayes rule.

Figures 8 and 9 show the motif assignments for two harbor regions. The gray level indicates the motif label. The consistency between the two assignments demonstrates that the proposed approach successfully learns the textures common to harbor objects. Subsequent analysis can use this knowledge to detect novel object instances in new remote sensed imagery to automatically update an object catalog (commonly known as a gazetteer). Such an approach is currently being investigated for the Alexandria Digital Library of geographic data at the University of California at Santa Barbara.²¹

5. Conclusion

We have described recent research into use of texture to analyze and manage large collections of remote sensed image and video data. A texture descriptor based on the outputs of scale and orientation-selective Gabor filters is shown to enable (1) content-based image retrieval in large collections of satellite imagery, (2) semantic labeling and layout retrieval in

an aerial video management system, and (3) statistical object modeling in geographic digital libraries.

This research is supported by the following grants: NASA California Space Grant (S. Newsam), U.S. Department of Transportation Research and Special Programs Administration contract DTRS-00-T-0002 (S. Newsam), U.S. Office of Naval Research N00014-01-1-0391 (L. Wang), National Science Foundation (NSF) Digital Library award IIS-9817432 (S. Bhagavathy), and NSF Instrumentation EIA-9986057.

References

1. A. Huertas and R. Nevatia, "Detecting buildings in aerial images," *Comput. Vision Graph. Image Process.* **41**(2), 131–152 (1988).
2. J. Shufelt and D. M. McKeown, "Fusion of monocular cues to detect man-made structures in aerial imagery," *Comput. Vision Graph. Image Process.* **57**(3), 307–330 (1993).
3. T. Chang and C.-C. J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Process.* **2**(4), 429–441 (1993).
4. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973).
5. F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: world features for image modeling and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(7), 722–733 (1996).
6. J. Mao and A. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recogn.* **25**(2), 173–188 (1992).
7. R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust. Speech Signal Process.* **33**, 959–963 (1985).
8. B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 837–842 (1996).
9. J. Daugman, "Complete discrete 2D Gabor transform by neural networks for image analysis and compression," *IEEE Trans. Acoust. Speech Signal Process.* **36**, 1169–1179 (1988).
10. S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Am.* **70**, 1297–1300 (1980).
11. P. Wu, B. S. Manjunath, S. Newsam, and H. D. Shin, "A texture descriptor for browsing and similarity retrieval," *J. Signal Process.* **16**(1–2), 33–43 (2000).
12. B. S. Manjunath, P. Salembier, and T. Sikora, eds., *Introduction to MPEG7: Multimedia Content Description Interface* (Wiley, New York, 2002).
13. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafine, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," *IEEE Comput.* **28**(9), 23–32 (1995).
14. J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, and C. F. Shu, "The Virage image search engine: an open framework for image management," in *Storage and Retrieval for Image and Video Databases IV*, I. K. Sethi and R. C. Jain, eds., *Proc. SPIE* **2670**, 76–87 (1996).
15. A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *Int. J. Comput. Vision* **18**(3), 233–254 (1996).
16. J. R. Smith and S.-F. Chang, "VisualSEEK: a fully automated content-based image query system," in *Proceedings of the Fourth ACM International Conference on Multimedia* (AMC Multimedia, Berkeley, Calif., 1996), pp. 87–98.
17. S. Newsam, J. Tešić, M. Saban, and B. S. Manjunath, MPEG-7

HomogeneousTextureDescriptorDemo,<http://vision.ece.ucsb.edu/texture/mpeg7/instructions.html>.

18. S. Z. Li, *Markov Random Field Modeling in Image Analysis* (Springer-Verlag, Berlin, 2001).
19. S. Bhagavathy, S. Newsam, and B. S. Manjunath, "Modeling object classes in aerial images using texture motifs," in *Proceedings of the International Conference on Pattern Recognition* (IEEE Computer Society, Los Alamitos, Calif., 2002), pp. 981–984.
20. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. R. Statistical Soc. Ser. B* **39**(1), 1–38 (1977).
21. The Alexandria Digital Library Project, <http://www.alexandria.ucsb.edu>.