Dynamic Learning of Shape for Automatic Object Recognition

Kyoung-Mi Lee
Computer Science Department
The University of Iowa
Iowa City, IA 52242 USA
klee1@cs.uiowa.edu

W. Nick Street
Management Sciences Department
The University of Iowa
Iowa City, IA 52242 USA
nick-street@uiowa.edu

Abstract

The problem of organizing and exploiting spatial knowledge for shape detection and recognition is an important task in many imaging domains, from medical diagnosis to automatic target recognition. In particular, clustering shapes allows for prototypes of each object type to be captured. This paper presents a new approach that uses unsupervised learning to find a set of templates specific to the objects being outlined by the user. The templates are formed by averaging the shapes that belong to a particular cluster, and allows the use of intelligent search to guide the order in which template matches are attempted. This results in decreased time and increased accuracy for repetitive segmentation problems, as system performance improves with continued use. Further, the information gained through clustering and user feedback is used to classify the objects for problems in which shape is relevant to the classification. The effectiveness of the resulting system is demonstrated on a medical diagnosis task using cytological images.

1. Introduction

The ultimate goal of computer vision is to simulate the human perception and interpretation of the world around us. Given an image of a scene, in terms of an array of pixels, the challenge is to locate and recognize different objects present in it. Shapes are important information for acquiring the notion of objects. There are many imaging applications where image analysis can be reduced to the analysis of shapes, (e.g. organs, cells, vehicles, machine parts, characters) (Hogg 1993; Katzir, Lindenbaum, & Porat 1994).

Usually an object recognition system using shape features consists of object extraction from the image, shape representation, and shape recognition. Our object recognition system constructs a prototype of a set of similar shape instances and learns to classify new instances based on feedback provided by a supervisor. A flowchart of the system is shown in Figure 1.

In order to extract an object in an image, early research in this area concentrated mainly on rigid shape matching to the model template, which can be recovered using correlation-based matching (Barnard & Thompson 1980) or the Hough transform (Hough 1962; Duda & Hart 1972). Because of the rigidness of the above approaches, their utility is limited. In most applications, an exact geometric model of the object is not available because of the variability in the imaging process and inherent within-class variabilities. We are interested in solving these problems with machine learning methods as a means for improving knowledge used in the imaging process, and thus for producing more robust software.

Much research has been done that utilizes shape learning to perform tasks in computer vision. Cho and Dunn used conjunctions of local shape properties and property-based learning (Cho & Dunn 1994). Bala and Wechsler used various morphological structuring elements and symbolic learning techniques (Bala & Wechsler 1996). Hsu and Hwang devised a method to generate descriptive classification rules of shape contours by using inductive learning (Hsu & Hwang 1997). Suganthan *et al.* presented a weighted attribute relational graph representation and used self organizing Hopfield network (Suganthan, Teoh, & Mital 1997). Amit *et al.* introduced a very large family of binary features for two-dimensional shapes to classify handwritten digits by using inductive learning during the construction of classification trees (Amit, Geman, & Wilder 1997). These results show that inductive machine learning techniques can significantly help the task of recognition patterns.

However, one drawback in existing research is that the characteristics extracted and learned by these systems are not easily understandable to humans. This is also the gap that exists between machine vision and human perception. Moreover, since most of them used nonincremental learning algorithms that require all the training examples at once, they have high computational complexity. Since our automatic object recognition algorithm must be performed on-line, efficiency is a basic requirement. Thus, we adopt an incremental approach that can benefit from solutions previously obtained on subset of objects. In this paper, we pro-

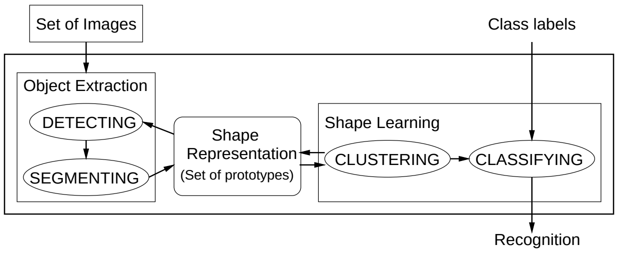


Figure 1: Flowchart of the proposed object recognition system

pose an incremental and dynamic approach to learning of shapes.

The remainder of the paper is organized as follows. Section 2 describes the first part of the proposed system in Figure 1, object extraction. Section 3 describes the shape representation for the system. Section 4 describes how the system discovers and classifies shapes using both supervised and unsupervised learning. Experimental results in Section 5 show that the quality of segmentation produced using the proposed algorithm is comparable to an exhaustive template search, and the computation time is significantly reduced. Conclusions are presented in Section 6.

2. Object Extraction

Object Detection: GHT

The Hough transform (HT) (Hough 1962) or its variant, the generalized Hough transform (GHT) (Ballard 1981) is a computationally efficient method of detecting or locating shapes. Generally, GHT requires a parameter space that is congruent to the image space, indexed by identical (x,y) parameters. For each edge pixel, an estimate of the location of a localization point, O, is made based on the gradient direction, θ, at the pixel. This estimation process produces a vector $\mathbf{v}$ which is used as a displacement from the edge pixel under observation. If the size and orientation of the sought-after shape is unknown, the computation involved in locating the shape increases drastically by adding dimensions to the parameter space. An iterative approach (IGHT) to GHT was proposed in (Oh 1997) which eliminates the extra dimensions by using a local accumulator (to find each template in turn) and a global accumulator (to collect the best scores from the local accumulator).

For each new image, IGHT is performed using the existing set of templates. After plateaus in the accumulator are removed with a peak-sharpening step, the highest value in the global accumulator is found. This represents the best match of any template to a shape in the image. The appropriate template is overlaid onto the image at the appropriate rotation angle. This process is repeated until a user-defined number of objects has been located.

An important aspect of the proposed method is that knowledge from shape learning allows the use of intelligent search to guide the order in which template matches are attempted. By searching first for the objects that we are most likely to find, we significantly reduce the expected time required, since not all templates will appear in every image. A user might also choose to search only for objects with a particular classification.

Segmentation: Snakes

To segment the exact boundaries of cell nuclei, the proposed system uses an adaptive spline curve-fitting technique, known as a snake (Kass, Witkin, & Terzopoulos 1988). Snakes use an energy minimizing spline guided by external constraint forces and influenced by image forces that pull it toward features such as lines and edges. In our case, the model is a closed curve that is attracted to strong edges in the image, and forms an arc in the absence of such edge information. The snakes are initialized using the results of IGHT that searches for ellipses of various sizes. Previous work (Oh 1997; Lee & Street 1999; Street 2000) has shown that the system isolates the cell nuclei very well using the combination of IGHT and snakes. The user can then edit the resulting outline by dragging the boundary points to their desired location. The user may also remove an incorrect boundary, or draw a boundary on an undetected object by hand, using the mouse to initialize the snake points. From the perspective of the learning method, this process creates a collection of positive examples of the shapes that this user would like to find and outline.

3. Shape Representation

The proposed system learns the boundary of templates used in the lookup table of GHT. Thus we need to find an appropriate shape model for incorporation into the GHT lookup table. Many different shape models have been proposed. Boundary decomposition is obtained by partitioning the shape boundary into curve segments (Liu & Srinath 1990). It is infeasible with GHT because the lookup table stores points, instead of curve segments. The Zahn and Roskies method used the tangent angle vs. arc length shape boundary representation (Zahn & Roskies 1972). Then the Fourier transform is applied to the boundary function and the resulting coefficients, invariant to scale change and translation, are used for shape description. Tangent angle vs. arc length representations suffer from very high noise sensitivity, and the Fourier transform requires the inverse function to get a raw points on boundary. Furthermore, if transform coefficients are processed, the inverse function may no longer produce accurate representations of the boundary. The chain code is an ordered sequence of critical points (Freeman 1974). In order to use this representation in learning, those points should be represented by the relationship with θ and they must be positioned consistently on each of the training examples such that a particular point always represents that same part of the shape on each example.

Shape using Centroid-radii

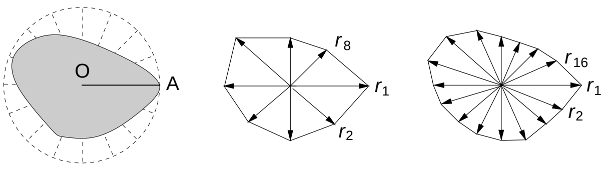
The essential components of the proposed system, a method of modeling the shapes of a set of variable objects, uses a Point Distribution Model (PDM) (Cootes *et al.* 1995). A PDM is a statistical shape model and is generated from a set of a training set of N object descriptions $(\mathbf{v}^k)^{k=1\cdots N}$. An object description, $\mathbf{v}^k$, is simply a labeled set of n points $(\mathbf{v}_i^k)_{i=1\cdots n}$. These positions must be positioned consistently on each of the training set.

We adopt the centroid-radii model to represent the shape by a set of points (r_{θ_i}) in polar coordinates (Chang, Hwang, & Buehrer 1991). Assume that OA is an arbitrary radius of the shape. In our algorithm, starting from OA and moving clockwise, we divide the circle into n equal arcs to place points around the boundary, with the regular interval being $\frac{360}{n}$ degrees. So the shape can be represented as a vector:

$$\mathbf{v} = (r_{\theta_1}, r_{\theta_2}, \cdots, r_{\theta_n})$$

where r_{θ_i}, $1 \leq i \leq n$, is the i-th radius from the centroid to the boundary of the shape and $\theta_i = \left(\frac{360}{n}\right) i$. The reflected shape can be represented as:

$$\mathbf{v}\text{'} = (r_{\theta_n}, r_{\theta_{n-1}}, \cdots, r_{\theta_1}).$$



(a) Object (b) n=8 (c) n=16

Figure 2: Shape representation

The two representations in Figure 2 ((b) and (c)) show this model can facilitate multiresolution representation. We wish to use only as many points as necessary to adequately model the shape.

A shape, $\mathbf{v}$, is considered similar to another, $\mathbf{u}$, if and only if the difference between $\mathbf{v}$ and $\mathbf{u}$ differs by a small value, ϵ. Formally, the difference between two shapes can be defined as follows:

$$\mathbf{d}_p(\mathbf{v}, \mathbf{u}) = \|\mathbf{v} - \mathbf{u}\|_p$$

where $\|\cdot\|_p$ is, e.g., the Manhattan (p=1), Euclidean (p=2), or max (p=∞) norm. In this application, we use the Manhattan norm. In other words, two shapes are considered similar if

$$\mathbf{d}(\mathbf{v}, \mathbf{u}) = \sum_{i=1}^{n} \left| r_i^{\mathbf{v}} - r_i^{\mathbf{u}} \right| < \epsilon$$

where $r_i^{\mathbf{v}}$ and $r_i^{\mathbf{u}}$ are radii in the two shapes. To avoid simple orientation differences, the comparison is performed at 16 different rotations.

Using this representation, a localization point in GHT becomes the center of shape. It is an easy and simple representation to understand and implement. Furthermore it is invariant to translation through the use of an object-centered coordinate system and to reflection through the use of the reflected shape. Since the algorithm is invariant to size and orientation, only one shape is needed. However, due to the lack of the internal information, we limit ourselves to topologically simple shapes with no holes.

4. Shape Learning

The incremental dynamic algorithm proposed in this section enables the use of the knowledge acquired from the previous step. Each time a new object is segmented, the algorithm evaluates whether to add it to an existing template or create a new template, without changing the existing templates. After choosing a template, it obtains the class name of the object to train the template classification.

Initialization

In this application the objects in question are human cell nuclei, which are more or less elliptical in shape. Therefore we begin by creating a set of elliptic templates with different shapes. Figure 3 shows some examples of elliptic boundaries. The initial templates are created automatically for this application but could easily be drawn by the user, or avoided completely, without changing the fundamentals of the algorithm.

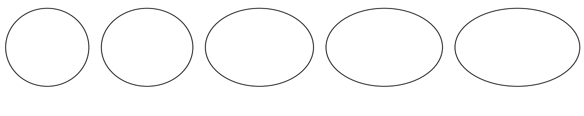


Figure 3: Examples of training templates

Unsupervised Shape Clustering

The objective of cluster analysis is to group a set of objects into clusters such that objects within the same cluster have a high degree of similarity, while objects belonging to different clusters have a high degree of dissimilarity. The structural relationships between individual data points, e.g. pronounced similarity of groups of data vectors, have to be detected in an unsupervised fashion.

The proposed clustering method proceeds as follows. If a segmented shape, $\mathbf{u}$, is similar one of existing templates, $\mathbf{t}$, the shape is added to the cluster represented by the templates. The first step is aligning the segmented example into the template. Given a set of N clustered shapes in the template, $\mathbf{v}^k = (r_i^k)_{i=1\cdots n}^{k=1\cdots N}$, the average shape, $\bar{\mathbf{v}}$, is calculated using

$$\bar{\mathbf{v}} = (\bar{\mathbf{v}}_i)_{i=1\cdots n} = \left(\frac{r_i^{\mathbf{u}} + \sum_{k=1}^{N} r_i^k}{N+1} \right)_{i=1\cdots n}.$$

If the shape is dissimilar with all existing templates, a new template is created with the shape of the new object. Thus, after the first image, each template represents a cluster of objects that are nearby one another in the space of possible shapes. The algorithm also records the scale of each matching object relative to the template. Figure 4 shows the clustering process graphically. The threshold that determines when a new shape has been found is set empirically, but is adjustable by the user.

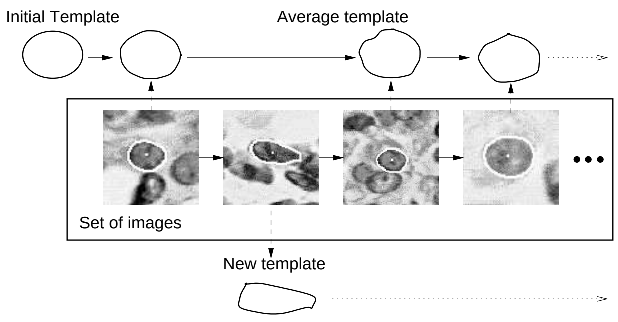


Figure 4: The clustering of shapes.

Classifying Templates

The ability to classify objects in an image plays an important role in the projected applications of this system. Consider the problem of isolating cell nuclei from heterogeneous tissue for dissection and genetic analysis. Such tissue may contain both diseased and healthy cells in the same sample. However, the molecular analysis may depend on collecting a "clean" sample of all diseased or all healthy cells for comparison purposes. Further, the user should not be expected to wait for the system to locate objects that are not desired for the particular experiment.

We take the approach that size and shape information already gathered in the clusters may be able to solve classification problems. The classification method proceeds as follows. As objects are isolated by the user, they are given a class label. A count is maintained of the number of objects from each class that are represented in each cluster. For instance, the third template shown in Figure 6 is made up of 39 benign and 35 malignant cell nuclei. These counts are stored based on the scale factor; for instance, Table 1 shows the scales and counts for the sixth template in Figure 6.

Table 1: Scale and diagnosis for the shapes in template 6

Scale	1.0	1.2	1.4	1.6	1.8
Ben/Mal Matches	3/0	16/0	0/3	0/0	0/1

As the training proceeds, the system is able to classify new objects based on the majority class of the matching template. This instance-based learning scheme was chosen for its simplicity and extensibility. Its primary function in this phase is to guide the template search. Thus it further improves the speed of the detection algorithm in cases where only a particular class of object is desired. For instance, if a user wants only malignant cells, the template search procedure will order the templates based on the probability of malignancy, rather than on the raw count of matched cells. Templates that match mostly benign cells will not be used until later in the search, even if their shapes were more common in previous images.

5. Experimental Results

The algorithm was tested on cytological images from fine-needle biopsies of breast masses, the same images used to train the original Xcyt system.[1] These images are gray-scaled with 640×480 pixel spatial resolution.

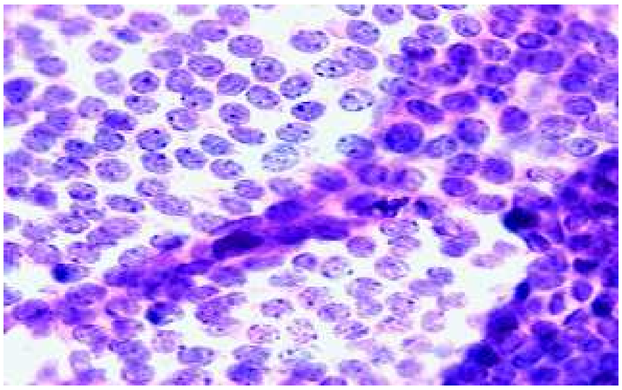


Figure 5: Sample image (reduced by 50%).

These images are classified as benign or malignant on a per-sample basis; no classification is available for

[1]Some of these images are available at `ftp://dollar.biz.uiowa.edu/pub/street/images/`.

individual cell nuclei. Therefore in these results we consider all nuclei in benign images to be benign and all nuclei in malignant images to be malignant. This assumption is reasonable but not entirely accurate, making the classification problem particularly difficult due to classification noise.

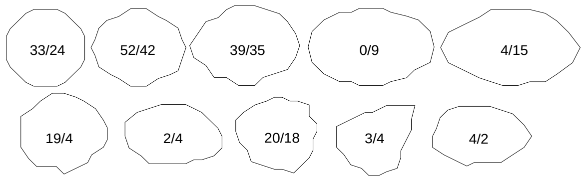


Figure 6: Learned template set

Figure 6 shows the set of templates constructed after training on six images. The first five shapes are modifications of the predefined templates in Figure 3. The following 5 new templates were created during the experiment. There are two numbers inside each template: the first one is the number of benign cells matching the template, and the second one is the number of malignant cells.

The test proceeded as follows. Two test images (one benign and one malignant) were set aside and used for all tests. The system was trained on a sequence of training images, alternating between benign and malignant, and tested on both test images after each training image.

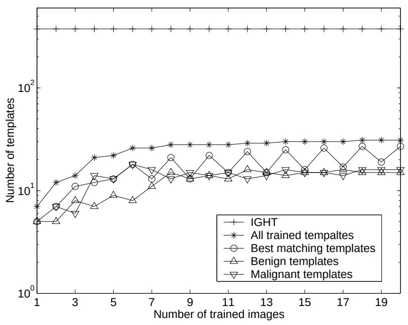


Figure 7: Number of templates searched

Figure 7 shows number of templates that had to be searched for the benign and malignant test images on a logarithmic scale. The search was carried out in four different ways. IGHT was performed with a constant set of elliptical templates. These vary significantly in size and shape in an attempt to capture the basic shape of all nuclei that might be encountered. The learning system builds a set of templates as it is used; the curve labeled "All trained templates" shows the number of these learned templates at each point. After an initial "ramping up" period, this number grows very slowly at a level nearly an order of magnitude below the original set of templates. The curve labelled "Best matching templates" shows the result of searching the templates in order of their frequency in previous images, and stopping when the accumulated frequency reached half of the total matching number. Since searching for a template with low frequency scarcely affects the accuracy, this results in a 50% reduction in the number of necessary templates, having a result equivalent in accuracy. Finally, the templates were ordered based on their classification, e.g., templates with the highest probability of being benign were searched first on the benign test image. This reduces the necessary search time even further, indicating that the instance-based classification method is able to distinguish reasonably well between nuclei types, even though the ground truth classifications in this problem were extremely noisy. This bodes well for future applications on heterogeneous tissue in which a trained expert will be available for the training phase.

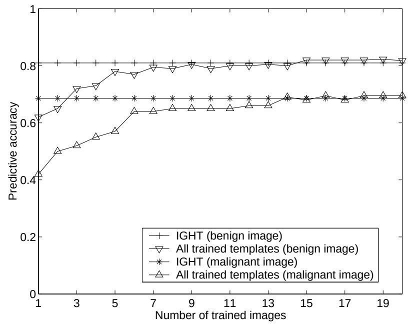


Figure 8: Predictive accuracy

Figure 8 shows the predictive accuracy of the segmentation method on the first 50 outlined nuclei (in the case of the benign image) or on all of the nuclei (in the case of the malignant image). As the system is trained on more and more images, its ability to correctly segment the nuclei in the test images increases. The accuracy of the new system reaches that of the exhaustive IGHT after about fifteen training images. In short, after a brief training phase, the new system has comparable accuracy to the original system, and achieves this performance in an order of magnitude less time per image.

Figure 9 shows the accuracy of the classification system in these tests. For each well-matched nuclei in a test image, the class is predicted based on the majority class of the template. After a few images, the classification rate stabilizes at 88% for the benign image and 65% for the malignant image.

6. Summary

In this paper, we have presented an incremental and dynamic learning algorithm to aid the location and

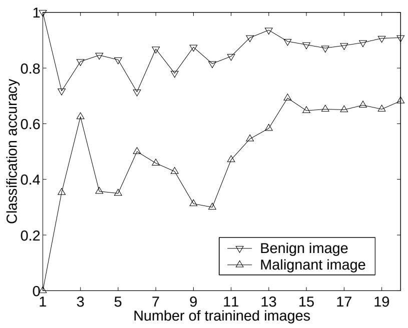


Figure 9: Classification accuracy

segmentation of objects in large-scale image analysis tasks. We have shown how our shape representation is applied in our image recognition system with the learning algorithm. The system demonstrates dramatic improvements in the time needed to perform template-matching search, without sacrificing the quality of the result. Future work will focus on the improvement of the instance-based classification algorithm for application to heterogeneous tissue, where the classification of different types of objects is of primary importance.

Acknowledgments

This work was partially supported by NIH grant CA64339-04 and NSF grant IIS-99-96044. The authors would like to thank Dr. William Wolberg of University of Wisconsin Department of Surgery for his helpful suggestions and for providing the test images.

References

Amit, Y.; Geman, D.; and Wilder, K. 1997. Joint induction of shape features and tree classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(11):1300–1305.

Bala, J., and Wechsler, H. 1996. Shape analysis using hybrid learning. *Pattern Recognition* 29(8):1323–1333.

Ballard, D. H. 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* 13(2):111–122.

Barnard, S. T., and Thompson, W. B. 1980. Disparity analysis of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2(4):333–340.

Chang, C. C.; Hwang, S. M.; and Buehrer, D. J. 1991. A shape recognition scheme based on relative distances of feature points from the centroid. *Pattern Recognition* 24(11):1053–1063.

Cho, K., and Dunn, S. M. 1994. Learning shape classes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(9):882–888.

Cootes, T. F.; Taylor, C. J.; Cooper, D. H.; and Graham, J. 1995. Active shape models-their training and application. *Computer Vision and Image Understanding* 61(1):38–59.

Duda, R. O., and Hart, P. E. 1972. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM* 15(1):11–15.

Freeman, H. 1974. Computer processing of line-drawing images. *ACM Computing Surveys* 6:57–97.

Hogg, D. S. 1993. Shape in machine vision. *Image and Vision Computing* 6(11):309–316.

Hough, P. V. C. 1962. Method and means for recognizing complex patterns. U.S. Patent 3069654.

Hsu, J. C., and Hwang, S. Y. 1997. A machine learning approach for acquiring descriptive classification rules of shape contours. *Pattern Recognition* 30(2):245–252.

Kass, M.; Witkin, A.; and Terzopoulos, D. 1988. Snakes: Active contour models. *International Journal of Computer Vision* 321–331.

Katzir, N.; Lindenbaum, M.; and Porat, M. 1994. Curve segmentation under partial occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(5):513–519.

Lee, K.-M., and Street, W. N. 1999. A fast and robust approach for automated segmentation of breast cancer nuclei. In *Proceedings of the IASTED International Conference on Computer Graphics and Imaging*, 42–47.

Liu, H., and Srinath, M. 1990. Partial shape classification using contour matching in distance transformation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12:1072–1079.

Oh, H. 1997. A memory-efficient generalized Hough transform for segmenting cytological images. Master's thesis, Oklahoma State University, Computer Science Department.

Street, W. N. 2000. Xcyt: A system for remote cytological diagnosis and prognosis of breast cancer. In Jain, L. C., ed., *Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis*. CRC Press. In press.

Suganthan, P. N.; Teoh, E. K.; and Mital, D. P. 1997. Optimal mapping of graph homomorphism onto self-organizing Hopfield network. *Image and Vision Computing* 15(9):679–694.

Zahn, C., and Roskies, R. 1972. Fourier descriptors for plane closed curves. *Computer Vision, Graphics, and Image Processing* 21:269–281.