Extraction of Skeletal Shape Features Using a Visual Attention Operator

Roman M. Palenichka, Rokia Missaoui, and Marek B. Zaremba

Dept. of Computer Science and Engineering
Université du Québec, Gatineau, Québec, Canada
{palenich,missaoui,zaremba}@uqo.ca

Abstract. The goal of the shape extraction method presented in this paper was to obtain a concise, robust, and invariant description of planar object shapes for object detection and identification purposes. The solution of this problem was chosen in the form of a piecewise-linear skeleton representation of local shapes in a limited number of salient object locations. A visual attention operator, which can measure the saliency level of image fragments, selects a set of most salient object locations for concise shape description. The proposed operator, called image relevance function, is a multi-scale non-linear matched filter, which takes local maxima at centers of locations of the objects of interest. This attention operator allows a simple extraction of vertices for the skeletal shape description by local maxima analysis.

1 Introduction

In a variety of image analysis tasks related to fast object detection and identification (verification), the main concern is adequate and concise representation of the object shape [1, 2]. The approach based on shape skeletons is, in the context of this application, efficient since it can represent in a very concise manner the topology of an object with several connected parts and shape details [3-6]. Such a description permits a complete morphological reconstruction of the planar shape provided local scale values (i.e., diameter values) are available in each skeleton point.

The classical skeletonization algorithms such as those based on an iterative (morphological) thinning and distance transformation provide the skeletal shape description but they are not robust to various shape distortions and noise [3, 4]. These methods are usually limited to process only binary images. Some multi-scale algorithmic generalizations to gray-scale images and three-dimensional (volume) images are also proposed [5-7]. Their performance strongly depends on the knowledge of some additional parameters, which are sensitive to distortions and irregularities. Complete skeletal shape is usually redundant to describe shape in the majority of object detection applications [1, 2]. Consequently, such a representation creates difficulties when comparing skeletal shapes, especially in the case of present noise and shape distortions.

More recently, several methods were developed to describe skeletal shapes in a piecewise-linear manner by skeleton vertices and their interconnections in the form of straight-line segments [8-10]. This is a concise representation of skeletal shapes with-

A. Fred et al. (Eds.): SSPR&SPR 2004, LNCS 3138, pp. 114–124, 2004.

out using classical skeletonization algorithms. For example, a statistical method of principal curves was used to extract directly the skeletal description of point sets [8, 9]. The algorithms for drawing principal curves using piecewise-linear approximation are, in their initial form, limited to simple curves or manifolds, where, for example, no intersections are allowed. Another kind of piecewise-linear skeletonization algorithms are based on unsupervised neural network methods, such as those based on self-organizing maps [10]. The shape skeleton can be obtained from a data-driven minimal spanning tree topology of a self-organizing map. The method is quite robust against sparse shapes and distortions but limited to process binary images and it deteriorates significantly if the segmented object contains components of various local sizes.

In this paper, we suggest a novel approach to skeletal shape description of gray-scale images based on the determination of salient object locations (i.e., interest points) and skeletal shape description relatively to the extracted locations. The whole object shape is described in terms of such local skeletal shapes and their relative positions and connectivity patterns. At the same time, this approach is an adaptation of the skeletal shape description for the case of object detection (localization) and identification in gray-scale images, without using an explicit image binarization. The development of the shape extraction method has the following objectives.

- Concise skeletal shape description by feature extraction in a selected number of salient object locations only.
- A simple distance (e.g., Euclidean distance) between shape feature vectors can compare two different shapes without computationally costly shape alignments.
- The shape features have to be invariant to geometrical transformations such as translation, scaling, and rotation.
- The method can process gray-scale images and have to be robust against noise and some local occlusions provided they do not occlude salient locations.

Salient object locations can be determined by the attention focusing approach, which was initially proposed to perform time-efficient search for objects of interest [11-13]. The underlying idea consists in focusing attention on the most salient image fragments or objects of interest, which are stable to intensity changes and shape geometrical transformations and can capture well the overall object shape. This is a biologically inspired approach that models basic elements of the visual perception and fast visual search in humans and animals. Given salient fragments, a complex object shape can be represented in terms of local shapes of these fragments and their relative positions on the image plane.

The determination of salient locations and shape feature extraction are both implemented by one visual attention operator called *image relevance function* (IRF) [14]. This operator is a multi-scale non-linear matched filter, which measures the saliency level of image locations and takes local maxima at the centers of locations of the objects of interest. The feature vector in each salient location includes local planar shape features and geometry features (i.e., parameters of affine transformations) such as relative position, local scale (size), and local orientation.

2 Morphological Modeling of Skeletal Shape Features

The proposed IRF method provides a description skeletal object shape in the form a set of most salient object locations each of them being described as a shape feature vector. The salient locations can be connected to each other provided connectivity conditions between two locations are fulfilled. The connected salient locations describe the whole (global) shape of a connected object. An image may contain many such connected objects each of them described by the local skeletal shape at salient locations. Moreover, each salient location can contain intensity (color) and texture features related to that location for object identification purposes using both local shape features and local intensity (texture) features.

In the IRF framework, planar shape features are separated and are independent from intensity features. Such a separation has a certain advantage over the integrated shape features extraction (e.g., features based on differentiation with Gaussian smoothing [11, 15]) because of the achieved invariance to transformation of translation, scaling and rotation and some intensity changes and lighting conditions. Additionally, a few intensity and texture features can be used for object intensity description to represent intensity variations as a texture, especially in the case of large scales (sizes of object regions).

For the purpose of multi-scale image analysis, a formal definition of a scale system is used [14]: a structuring element at scale n of a *uniform scales system* is formed by the morphological dilation (denoted by $\oplus$) by S_0, $S_n = S_{n-1} \oplus S_0$, n=1, 2,...,M-1, where M is the total number of scales and the structuring element S_0 defines the minimal scale and object resolution. The structuring elements have the same shape such as the disk shape (see examples in Fig. 1). The above is a morphological definition of scales, which is different from the notion of scales in the scale-space filtering [11].

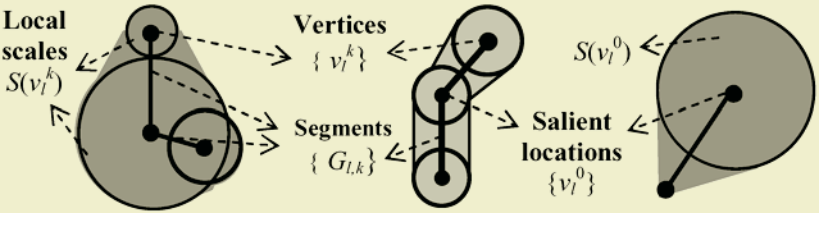

Fig. 1. Examples of local skeletal shape features.

We have proposed a piecewise-linear local skeletal description of planar shapes related to salient object locations. This skeletal shape representation is an economical approach to shape description. An object local shape is related to a particular salient location v_l^0 and the local scale value at that location, $S(v_l^0)$ (see Fig. 1). Given K vertices and K scale values associated with each vertex, the local planar shape (as a support region U) of an object of interest located at v_l^0 is formed by the dilation operations of skeleton straight-line segments $\{G_{l,k}\}$ with size-variable structuring elements, $\{S(G_{l,k})\}$:

$$U(v_l^0) = \bigcup_{k=1}^{K} G_{l,k} \oplus S(G_{l,k}) = \bigcup_{k=1}^{K} \bigcup_{v_l^m \in G_{l,k}} S(v_l^m), \tag{2.1}$$

where $\oplus$ denotes the morphological dilation, $S(v_l^m)$ is a structuring element with variable size (e.g., diameter r_m) as a function of point $v_l^m \in G_{l,k}$, and K is the maximal topological order of the skeleton vertices. The value of diameter r_m is a linear combination of the scale sizes r_0 and r_k at terminal vertices v_l^0 and v_l^k of segment $G_{l,k}$. Equation (2.1) represents a method of *scale-interpolated dilation* in the piecewise-linear modeling of skeletal shapes and

$$r_m = \frac{d(v_l^k, v_l^m)}{d(v_l^k, v_l^0)} \cdot r_0 + \frac{d(v_l^0, v_l^m)}{d(v_l^k, v_l^0)} \cdot r_k, \tag{2.2}$$

where $d(.,.)$ is the Euclidean distance between two skeleton vertices on the image plane.

The whole planar shape of a multi-scale object of interest is formed by pair-wise concatenations of the local shapes at L vertices $\{v_l^0, l=1,...,L\}$ if the connectivity between the corresponding vertices can be established. For each l, the local skeleton vertices $\{v_l^k, k=1,...,K\}$ can be considered as shape details at that vertex, i.e., respective salient location.

This model of planar local shape is associated with an intensity model of image fragment of size $2r_l$ centered at v_l^0. The intensity modeling involves two dominant intensity levels with an additive noise model, which can also represent a textured intensity, in order to descrobe image intensity locally and concisely [14].

3 Determination of Salient Locations Using Image Relevance Function

Each salient object location is associated with its own salient fragment centered at a particular local maximum of the IRF. The IRF is defined generically as an image operator, which takes local maximal values at centers of salient image fragments and can be used to describe objects of interest in the salient locations. At certain conditions, the IRF maximums are positioned on object medial axes or at the centers of its parts, which are relevant to shape description (see Fig. 2). In order to address the aforementioned problems of skeletal shape extraction for object detection it is suggested to apply an improved version of a model-based IRF described in the context of object detection [14]. Localization of salient image fragments is based on a fast computation of the multi-scale IRF and determination of its local maxima. The positions of local maximum values of the multi-scale IRF coincide with location points of the salient image objects in a region of interest A:

$$(i_f, j_f)_l = \max_{(i,j) \in A} \max_k \{\Phi[(g(i,j), S_k], (i,j) \notin \Gamma_l\}, \tag{3.1}$$

where $g(i,j)$ is the input gray-scale image, $\Phi[g(i,j),S_k]$ is a non-linear matched filter at kth scale, and $(i_f,j_f)_l$ are two coordinates of lth maximum. The region $\Gamma_l \subset A$ corresponds to the masking region, which excludes determined maximum points from further analysis.

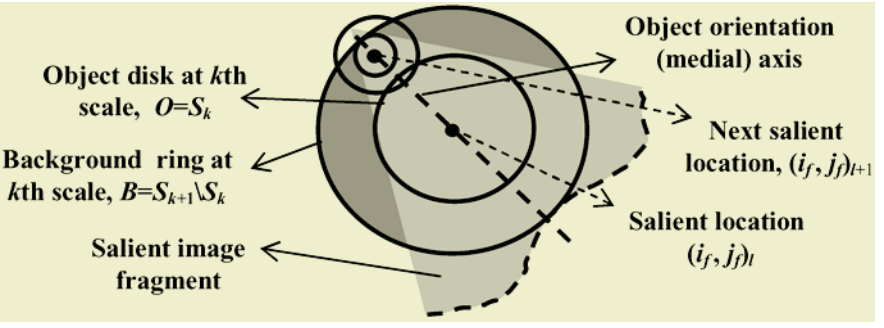

Fig. 2. Illustration of the relevance function computation for the single-scale case.

Four saliency conditions are considered in the design of $\Phi(g(i,j),S_k)$: 1) significant local contrast; 2) local homogeneity of object intensity; 3) specific object intensity range; 4) specific range of object sizes and shape of the scales $\{S_k\}$. The first condition is described by the absolute value for the local object-to-background contrast. The local homogeneity condition means that the intensity variance is relatively small in the object region. The intensity range means specific values for the object intensity in order to distinguish it from the background or other objects. Since the measures for contrast, homogeneity and intensity range involve object disk regions and background ring regions of a particular range, the IRF will take implicitly into account shape and scale constraints (condition) of the objects. Taken these conditions, the IRF can be computed in point (i,j), at scale S_k as follows:

$$\Phi[g(i,j),S_k] = c^2(i,j,S_k) - \alpha \cdot d^2(i,j,S_k) - \beta \cdot e^2(i,j,S_k), \tag{3.2}$$

where $c(i,j,S_k)$ is an estimate for the local contrast, $d(i,j,S_k)$ is an estimate for intensity deviation in the object region, $e(i,j,S_k)$ is the object intensity shift, α and β are constraint coefficients which control the contributions of the two constraints to the overall value of IRF. An estimate of the optimal value of α and β in the sense of the maximum likelihood decision can be computed assuming some distributions (e.g., Gaussian functions with different parameters) for the three variables in Eq.(3.2) under the condition of object presence in point (i,j). The constraint coefficients are inversely proportional to the variances of two constraints in the case of Gaussian distributions: $\alpha = \sigma_c^2 / \sigma_d^2$ and $\beta = \sigma_c^2 / \sigma_e^2$. For example, the contrast estimate $c(i,j,S_k)$ is the intensity difference,

$$c(i,j,S_k) = f_1(i,j,S_k) - f_0(i,j,Q_k), \tag{3.3}$$

where $Q_k = S_{k+1}/S_k$, is the background estimation region at scale k, i.e., a ring around the disk S_k. $f_1(i,j,S_k)$ and $f_0(i,j,Q_k)$ are the mean values of $g(i,j)$ in regions S_k and Q_k, respectively (see Fig. 2). The mean square deviation was used for the estimation of $d(i,j,S_k)$ in Eq.(3.2). The object intensity shift is measured as a deviation of the mean intensity value $f_1(i,j,S_k)$ from an object intensity of reference.

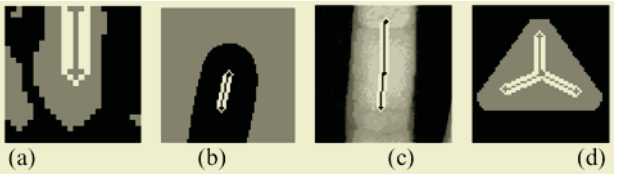

Fig. 3. Examples of local shape features (piecewise-linear skeletons) extracted at salient image locations.

4 Extraction of Skeletal Shape Features

A so-called *saliency hypothesis* is tested first in each local maximum point before the shape feature extraction. It consists of comparisons of local contrast and local homogeneity with saliency thresholds [14].

The extraction of skeletal shape features uses mostly intermediate results of IRF computation, (Eq. 3.2), and is computationally insignificant as compared to the IRF calculation. The invariance parameters for the considered geometrical transformations (translation, scaling, and rotation) are computed with respect to the current local maximum of the IRF. The first parameter is the absolute position of the lth salient location, v_l^0, consisting of two coordinates $(i_f,j_f)_l$. The next two parameters, local scale and local orientation, which are related to point $(i_f,j_f)_l$, are estimated using intermediate results of the IRF calculation (see Fig. 2). The local scale is determined by the contrast maximization,

$$\rho(i_f, j_f) = \arg\max_k \left\{ c^2(i_f, j_f, S_k) - \alpha \cdot d^2(i_f, j_f, S_k) \right\} \tag{4.1}$$

where the variables and the constant coefficient α have the same meanings as in Eq.(3.1) and Eq.(3.2). Object orientation can be estimated in a simple way since the next maximum point $(i_f,j_f)_{l+1}$ in the current region of attention with respect to the focus of attention $(i_f,j_f)_l$ provides the orientation vector (see Fig. 2).

The proposed IRF approach provides at the same time a simple method to determine vertices for the piecewise-linear skeletal representation of object local shapes in salient locations. This can be done by analysing consecutive K maximums of IRF next to a given salient location v_l^0. Such a procedure determines K local skeleton vertices $\{v_l^k, k=1,\ldots,K\}$, which all are connected to vertex v_l^0 according to the morphological model in Eq. (2.1). Given a neighborhood region $B(v_l^0)$ around vertex v_l^0, the algorithm for the local shape feature extraction is as follows if starting from k=1.

Step 1. Determine location v_l^k of the kth local maximum of IRF in the masked neighborhood region $B(v_l^0)$, non-including previous (k-1) IRF maxima.

Step 2. Test the saliency hypothesis with respect to kth local maximum point. If the testing outcome is positive then go to *Step* 3, otherwise go to *Step* 4.

Step 3. Determine relative scale and relative orientation associated with v_l^k. Attach these values to the shape vector at positions $2k$ and $(2k+1)$. If $k<K$ then increment $k=k+1$, and mask the neighborhood of vertex v_1^k and go to Step 1, otherwise go to *Step* 4.

Step 4. Attach two coordinates of the vertex v_l^0, its local scale and orientation at the first four positions of the local shape vector. If $k<K$ then the remaining $2\cdot(K-k)$ shape features are set to zero.

The accuracy of shape features determined by this algorithm depends on the correspondence of processed images to the underlying model. Examples of detected salient image fragments in real images with superimposed skeleton fragments are shown in Fig. 3. For the purpose of object verification, intensity and texture features can also be attached to the shape feature vector. They may include object mean intensity, color intensity components, local contrast, and local object variance. The texture features can be used as well in order to describe concisely intensity fluctuations for large scales (object sizes) depending on the application.

Determined salient locations, $\{v_l^0,\ l=1,...,L\}$, with extracted shape features can be connected to each other if the connectivity between them can be established. This will provide a complete and two-level skeletal shape description of the objects of interest. We were looking for the connectivity in the form of a *spanning tree* (*forest*) constructed in an optimal way. The connectivity algorithm is based on the Markov random chain models of vertex connectivity and finding optimal connectivity pattern between all the salient location vertices $\{v_l^0, l=1,...,L\}$ by the maximization of joint probability $P_c(v_1^0,\ldots,v_L^0)$ of vertex connectivity:

$$P_c(v_1^0,...,v_L^0) = \prod_{v_s^0,v_s^0,v_t^0} P_c(v_q^0 \wedge v_s^0 \,/\, v_s^0 \wedge v_t^0) \quad \Rightarrow \quad \max_{v_s^0,v_s^0,v_t^0}, \tag{4.2}$$

where $P_c(v_q^0 \wedge v_s^0 / v_s^0 \wedge v_t^0)$ is the probability that skeleton vertices v_q^0 and v_s^0 are connected with each other provided the precedent vertex v_t^0 in the spanning tree is connected to v_s^0.

The probabilistic connectivity framework was chosen because it provides an optimal solution in the case of gray-scale images, when object shape is sparse with noise presence and possible local occlusions. At certain model assumptions (or particular likelihood functions) this problem can be reduced to the problem of finding minimum spanning tree for a set of extracted vertices, $\{v_l^0,\ l=1,...,L\}$. The likelihood function, $P_c(v_q^0 \wedge v_s^0,\ a_{q,s},\ d_{q,s} / v_s^0 \wedge v_t^0)$, which is involved in the probability model in Eq. (4.2), uses intensity values $\{a_{q,s}\}$ and distances $\{d_{q,s}\}$ between the vertices as the connectivity variables. The complete description of this algorithm is out of scope of this paper and some details for global skeletal shape extraction can be found in Ref. [10, 16].

Table 1. Measured accuracy of shape feature extraction (error given in pixel resolution) versus contrast-to-noise ratio.

Contrast-to-noise ratio	2	4	8	16	32
Localization error	1.4	1.1	0.7	0	0
Scale error	3.2	2.9	0.9	0.1	0
Orientation error	2.4	1.2	0.3	0.4	0

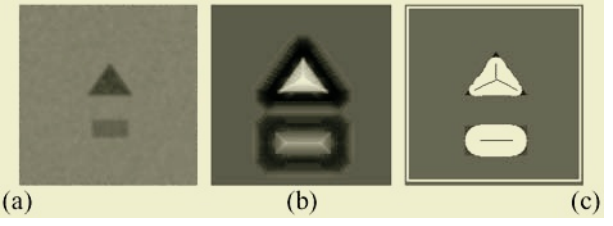

Fig. 4. Illustrative example of IRF calculation (b) and local shape extraction (c) using six most salient IRF maximum points as applied to initial synthetic image in (a).

5 Experimental Results

The first kind of experiments was the performance testing of the IRF approach to shape feature extraction based on synthetic images with known values of the shape features. The position (two coordinates), scale, and orientation values of the salient locations determined by the proposed IRF have been measured and compared with the reference values to determine the accuracy. Image noise has been imitated in synthetic images in order to calculate the accuracy as a function of ratio of the object-to-background contrast and noise magnitude (standard deviation). An example of used synthetic image objects with known shape features and added noise is shown in Fig. 4. The results of accuracy testing are given in Table 1. The error in feature values was measured in pixel resolution relatively to the correct feature values. In particular, the scale error was measured in pixels as the deviation of the scale diameter. The orientation error was measured in term of the displacement of the second most salient maximum of IRF with respect to its correct position. Analysis of these data shows good accuracy and robustness of the proposed approach to feature extraction.

The objective of the second kind of experiments was the visual evaluation of the IRF performance in extracting skeletal shape in application to biometrical and medical imaging. One example of using IRF approach to detect fingers and determine their shape for the purpose of a biometrical identification from a hand image is shown in Fig. 5. This is an example of vertex extraction by IRF local maxima and establishing piecewise-linear connectivity between the extracted vertices. A detail analysis of finger geometry and texture of the finger skin have to be performed in each salient location. The skeletal shape was extracted directly from the grey-scale image in Fig. 5a without the image binarization.

The proposed IRF method was also compared with the skeletonization method using self-organizing maps by applying both methods to the same test image [10]. An example of obtained results by the two methods is shown in Fig. 6. The method of piecewise-linear skeletonization using self-organizing maps performed worse even

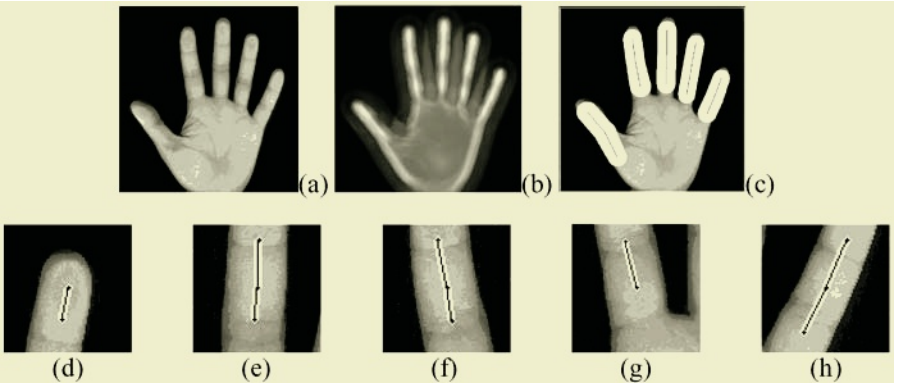

Fig. 5. Results of skeletal shape extraction of fingers obtained directly from the gray-scale image of a hand: (a) – initial image; (b) – IRF calculation; (c) – skeletal shape of fingers. Examples of local shapes at salient locations are shown below.

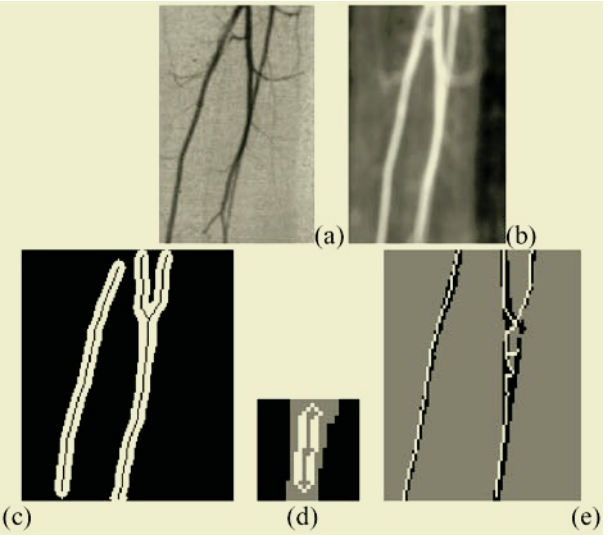

Fig. 6. An illustration to skeletal shape extraction in digital angiography: input image, (a); image of IRF, (b); skeletal shape of main blood vessels in the selected region of interest, (c); most salient object fragment, (d); result of skeletonization using method of self-organizing maps [10], (e).

when applied to the binary version of the input image and gave visible imprecision such as jaggedness of lines.

6 Conclusions

A method for the extraction of skeletal shape features using a visual attention operator was developed. It is based on the determination of salient object locations by local

maxima analysis of the introduced multi-scale IRF. The same IRF approach was applied to extract a piecewise-linear skeletal shape at determined salient locations. The proposed concise description of local shapes has the following advantages in the context of object detection and shape verification. The shape comparison does not require computationally complex alignments because two different shapes can be compared by a simple distance measure (e.g., Euclidean distance). The IRF approach provides a robust shape extraction directly from gray-scale images, in the presence of noise and under some local distortions. The obtained shape features can easily become invariant with respect to translation, scaling, and rotation by a normalization relatively to geometrical parameters for a current location.

Acknowledgments

We are grateful to VRQ (Valorisation Recherche Québec) and Canadian Heritage for their financial support to CoRIMedia (Consortium de Recherche en Image et Multimedia).

References

1. M. D. Wheeler and K. Ikeuchi, “Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition”, *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 17, No. 3, pp. 252-265, 1995.
2. V. Conception and H. Wechsler, “Detection and localization of objects in time-varying imagery using attention, representation and memory pyramids”, *Pattern Recognition*, Vol. 29, No. 9, pp. 1543-1557, 1996.
3. N. Blum and R.N. Nagel, “Shape description using weighted symmetric axis features”, *Pattern Recognition*, Vol. 10, pp. 167-180, 1978.
4. Y. S. Chen, and Y.T. Yu, “Thinning approaches for noisy digital patterns”. *Pattern Recognition*, Lol. 29, No. 11, pp. 1847-1862, 1996.
5. G. Borgefors, “Distance transformation in digital images”, *Vision, Graphics, and Image Processing,* Vol. 34, pp. 344-371, 1986.
6. G. Borgefors, G. Ramella, G. Sanniti di Baja, and S. Svenson, “On the multi-scale representation of 2D and 3D shapes”, *Graphical Models and Image Processing,* Vol. 61, pp. 44-62, 1999.
7. C. Archelli and G. Ramella, “Sketching a grey-tone pattern from its distance transform,” *Pattern Recognition,* Vol. 29, No. 12, pp. 2033-2045, 1996.
8. T. Hastie, and W. Stuetzle, “Principal curves”, *Journal of the American Statistical Association*, Vol. 84 (406), pp. 502-516, 1989.
9. B. Kegl, *et al.*, “Learning and design of principal curves”, *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 22, No. 3, pp. 281-297, 2000.
10. R. Singh, V. Cherkassky, and N. Papanikopoulos, “Self-organizing maps for the skeletonization of sparse shapes”, *IEEE Trans. on Neural Networks*, Vol. 11, No. 1, pp. 241-248, 2000.
11. T. Lindeberg, “Detecting salient blob-like image structures and their scale with a scale-space primal sketch: a method for focus of attention”, *Int. Journal of Computer Vision*, Vol. 11, pp. 283-318, 1993.

12. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 20, No. 11, pp. 1254-1259, 1998.
13. H. D. Tagare, K. Toyama, and J.G. Wang, "A maximum-likelihood strategy for directing attention during visual search", *IEEE Trans. Pattern Analysis and Machine Intelligence,* Vol. 23, No. 5, pp. 490-500, 2001.
14. R. M. Palenichka, "A visual attention operator based on morphological models of images and maximum likelihood decision", *Proc. Int. Workshop SSPR 2002, LNCS* 2396, pp. 310-319, 2002.
15. J. J. Koenderink and A. J. van Doorm, "Representation of local geometry in the visual system", *Biological cybernetics,* Vol. 55, pp. 367-375, 1987.
16. M. B. Zaremba and R. M. Palenichka, "Probabilistic morphological modeling of hydrographic networks from satellite imagery using self-organizing maps," *Control & Cybernetics*, Vol. 31, No. 2, pp. 343-370, 2002.