Affine-Invariant Curve Normalization for Shape-Based Retrieval

Yannis Avrithis, Yiannis Xirouhakis and Stefanos Kollias

Image, Video and Multimedia Systems Laboratory
Department of Electrical and Computer Engineering
National Technical University of Athens
E-mail: {iavr,jxiro}@image.ntua.gr

Abstract

A novel method for two-dimensional curve normalization with respect to affine transformations is presented in this paper, allowing an affine-invariant curve representation to be obtained without any actual loss of information on the original curve. It can be applied as a pre-processing step to any shape representation, classification, recognition or retrieval technique, since it effectively decouples the problem of affine-invariant description from feature extraction and pattern matching. Curves estimated from object contours are first modeled by cubic B-splines and then normalized in several steps in order to eliminate translation, scaling, skew, starting point, rotation and reflection transformations, based on a combination of curve features including moments and Fourier descriptors.

1. Introduction

Due to the recent growth of interest in multimedia applications, an increasing demand has emerged for efficient storage, management, browsing and content-based retrieval in multimedia databases [1]. In this context, content information in is usually modeled in terms of low-level features such as color, texture, motion and shape. As observed in [7], it is natural to retain only the object boundary if it contains the main information for description – for example, in classification of silhouettes of airplanes, or in character recognition. The study of shape for general object description is currently an active field of research, mainly because (i) shape can provide a powerful tool for visual image retrieval, by means of a query-by-sketch mechanism [3], and (ii) content-based functionalities will be embedded in new multimedia coding standards [5], employing second-generation, shape-based coding techniques [6].

Several methods have been proposed in the literature for shape analysis, modeling and representation, ranging from chain coding and polygonal approximation to B-splines [4] and Fourier descriptors [7]. Most approaches face the problem of drastic shape changes due to viewpoint (perspective) transformations, which can be well approximated by affine transformations when objects are far from the camera. Several *affine invariant* techniques have thus been proposed; in most of them, however, invariance is 'embedded' in the process of matching, recognition, or similarity measure estimation. For example, affine invariant shape features and similarity measures can be employed [8], with the main disadvantage that most information about the original curve is lost. An alternative is to match two given curves by optimally evaluating their affine parameters [7], with the disadvantage of high computational cost and the requirement of *a priori* knowledge of both shape instances.

To this end, a novel method for two-dimensional curve *normalization* with respect to affine transformations is presented, making it possible to obtain an affine-invariant curve representation without any actual loss of information. Curve shapes are normalized to a 'standard' position, defined in such a way that all affine transformations of the same object are also normalized to the same position; apart from the affine transformation parameters, no other information is discarded. Hence, the proposed method can be applied as a pre-processing step to any shape representation, classification or recognition technique (e.g., shape matching using deformable templates [3]), since it decouples affine-invariant description from feature extraction and pattern matching.

In particular, a closed curve representing the contour shape of an object is first modeled by a cubic B-spline so that the shape is simplified, segmentation noise is reduced, and uniform curve sampling in terms of arc length is obtained. Then, the sampled curve is normalized in several steps in order to eliminate translation, scaling, skew, starting point, rotation and reflection transformations. Normalization is based on a combination of curve features including moments and Fourier descriptors. All such features are globally estimated from all curve samples; no local information is used. The computational complexity involved is negligible, so that the method can be easily integrated in a real-time image retrieval or video coding system.

2. B-Spline Curve Modeling

In the following, it is assumed that the contour shape of an object is available and represented by a set of ordered points forming a two-dimensional, planar and closed curve obtained from image data by means of manual or automatic segmentation. The M-RSST color segmentation algorithm [1] was actually used in our experiments, combined with motion segmentation in the case of video sequences. Since discretization is involved in the segmentation process, leading to segmentation noise and non-uniform sampling in terms of arc length, a *B-spline* curve model is employed. B-splines have been widely employed for shape analysis and modeling, since they possess a number of important properties such as smoothness and continuity, built-in boundedness and local controllability [4].

Cubic B-splines are composite curves consisting of a large generally number of connected curve segments with C^2 continuity on the connection points. Each segment is a linear combination of four cubic polynomials, commonly known as *basis functions*. The whole B-spline is characterized by a number of *control points*, equal to the number of curve segments. Given a set of data curve points, obtained from segmentation, the control points are determined by fitting the B-spline to the data points in a MMSE sense. Since in general different sets of control points may describe the same curve, the *knot points*, defined as the connection points between curve segments, are derived from a linear combination of the estimated control points. Finally, in order to achieve uniform sampling in terms of arc length, an appropriate parametric value is obtained using the *chord length* (CL) method [4], based on the fact that the chord length between any two points is a very close approximation to

the arc length; knot points are re-allocated with equal spacing in terms of the estimated parametric value.

3. Curve Orthogonalization

A *curve orthogonalization* procedure is employed as the first stage of normalization, effectively normalizing a curve with respect to possible translation, skew and scaling and reducing affine transformations to orthogonal ones. Let $\mathbf{s}_i = [x_i \ y_i]^T$, $i = 0,1,\ldots,$ N-1, be N curve points obtained through B-spline modeling. A $2 \times N$ matrix notation $\mathbf{s} = [\mathbf{s}_0 \ \mathbf{s}_1 \ \ldots \ \mathbf{s}_{N-1}]$ is used to represent the points, while their horizontal and vertical coordinates are represented by $\mathbf{x} = [x_0 \ x_1 \ \ldots \ x_{N-1}]$ and $\mathbf{y} = [y_0 \ y_1 \ \ldots \ y_{N-1}]$. The (p,q)-order moments

$$m_{pq}(\mathbf{s}) = \frac{1}{N}\sum_{i=0}^{N-1} x_i^p y_i^q \qquad (1)$$

of order up to 2 are used for the construction of the *normalized curve* $n_a(\mathbf{s})$. The procedure comprises a set of linear operations – translation, scaling and rotation. First, the center-of-gravity of the curve is normalized so as to coincide with the origin:

$$\mathbf{x}_1 = \mathbf{x} - \mu_x, \quad \mathbf{y}_1 = \mathbf{y} - \mu_y \qquad (2)$$

where $\mu_x = m_{10}(\mathbf{s})$, $\mu_y = m_{01}(\mathbf{s})$. The curve is then scaled so that its 2nd-order moments become equal to 1:

$$\mathbf{x}_2 = \sigma_x \mathbf{x}_1, \quad \mathbf{y}_2 = \sigma_y \mathbf{y}_1 \qquad (3)$$

where $\sigma_x = 1/\sqrt{m_{20}(\mathbf{s}_1)}$, $\sigma_y = 1/\sqrt{m_{02}(\mathbf{s}_1)}$. A counter-clockwise rotation by $\theta_0 = \pi/4$ follows:

$$\mathbf{s}_3 = \mathbf{R}_{\pi/4}\mathbf{s}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{x}_2 - \mathbf{y}_2 \\ \mathbf{x}_2 + \mathbf{y}_2 \end{bmatrix} \qquad (4)$$

Finally, the curve is scaled again, exactly as in (3):

$$\mathbf{x}_4 = \tau_x \mathbf{x}_3, \quad \mathbf{y}_4 = \tau_y \mathbf{y}_3 \qquad (5)$$

where $\tau_x = 1/\sqrt{m_{20}(\mathbf{s}_3)}$, $\tau_y = 1/\sqrt{m_{02}(\mathbf{s}_3)}$. The normalized curve $n_a(\mathbf{s}) \equiv \mathbf{s}_4$ can also be written as

$$n_a(\mathbf{s}) = \mathbf{N}(\mathbf{s})(\mathbf{s} - \mu(\mathbf{s})) \qquad (6)$$

where $\mu(\mathbf{s}) = [m_{10}(\mathbf{s}) \ m_{01}(\mathbf{s})]^T$ and $\mathbf{N}(\mathbf{s})$ denotes the 2×2 *normalization matrix* of $\mathbf{s}$. It can proved [2] that the normalized curve $n_a(\mathbf{s})$ always has the following properties:

$$m_{10}(n_a(\mathbf{s})) = m_{01}(n_a(\mathbf{s})) = m_{11}(n_a(\mathbf{s})) = 0 \qquad (7a)$$

$$m_{20}(n_a(\mathbf{s})) = m_{02}(n_a(\mathbf{s})) = 1 \qquad (7b)$$

Moreover, the above conditions can only be achieved if the rotation angle used in (4) is equal to $k\pi/2 + \pi/4$, $k \in \mathbb{Z}$. The term *orthogonalization* is justified since these conditions are equivalent to $n_a(\mathbf{s})(n_a(\mathbf{s}))^T = \mathbf{I}$. Let us now consider two curves $\mathbf{s}$, $\mathbf{s}'$ related through an affine transformation:

$$\mathbf{s}' = \mathbf{A}\mathbf{s} + \mathbf{t} = \begin{bmatrix} \mathbf{x}' \\ \mathbf{y}' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \qquad (8)$$

where matrix $\mathbf{A}$ is assumed to be of full rank. Then, $\mathbf{s}'_1 = \mathbf{s}' - \mu(\mathbf{s}') = \mathbf{A}(\mathbf{s} - \mu(\mathbf{s})) = \mathbf{A}\mathbf{s}_1$ and translation is removed. Moreover, when a normalized curve is rotated or reflected, in which case $\mathbf{A}$ is orthogonal, it remains normalized. Inversely, if both curves are normalized, then $\mathbf{A}$ should be orthogonal. It is thus shown in [2] that there exists an orthogonal 2×2 matrix $\mathbf{Q}$ such that

$$n_a(\mathbf{s}') = \mathbf{Q}\,n_a(\mathbf{s}) \qquad (9)$$

Affine transformations are reduced to orthogonal ones that may contain only rotation and/or reflection; hence, normalized curves are invariant to translation, scaling and skew transformations. Note that normalization is performed *without knowledge* of the affine parameters $\mathbf{A}$, $\mathbf{t}$, and *without one-to-one matching* between curves $\mathbf{s}$ and $\mathbf{s}'$. Furthermore, it can be seen that the set of transformation parameters $\{\mu_x, \mu_y, \sigma_x, \sigma_y, \tau_x, \tau_y\}$ along with $n_a(\mathbf{s})$ contain all information about the original curve $\mathbf{s}$.

4. Starting Point Normalization

Starting point normalization for closed curves is applied at this point; rotation and reflection normalization follows since it depends on the starting point. Both normalizations are based on the discrete Fourier transform instead of curve moments. The complex vector notation $\mathbf{z} = \mathbf{x} + j\mathbf{y} = [z_0 \ z_1 \ \ldots \ z_{N-1}]^T$ is thus used for curve representation, where $z_i = x_i + jy_i$, $i = 0,1,\ldots,N$-1, denotes a single curve point. The DFT $\mathbf{u} = \Im(\mathbf{z})$ of curve $\mathbf{z}$ is given by

$$u_k = \sum_{i=0}^{N-1} z_i w^{-ki}, \quad k = 0,1,\ldots,N-1 \qquad (10)$$

where $w = e^{j2\pi/N}$, so that $w^{\ell N} = 1$, $\ell \in \mathbb{Z}$. Employing the *primary argument*, or *phase* $a_k = \operatorname{Arg} u_k$ we construct the *phase vector* $\mathbf{a} = \operatorname{Arg}\mathbf{u} = [a_0 \ a_1 \cdots a_{N-1}]$. Consider now a second curve $\mathbf{z}' = [z'_0 \ z'_1 \cdots z'_{N-1}]$ that is circularly shifted with respect to $\mathbf{z}$ by m samples, where $m \in \{0,1,\ldots,N$-1$\}$:

$$\mathbf{z}' = S_m(\mathbf{z}): \quad z'_i = z_{(i+m)\bmod N}, \quad i = 0,1,\ldots,N-1 \qquad (11)$$

Then it can be shown that $u'_k = w^{km}u_k$, $k = 0,1,\ldots,N$-1, or

$$a'_k = (a_k + 2\pi km/N) \bmod 2\pi, \quad k = 0,1,\ldots,N-1 \qquad (12)$$

Based on this property, we define a *standard* circular shift using the first and last Fourier phases:

$$p(\mathbf{z}) = \left\lfloor \frac{N}{4\pi}(a_1 - a_{N-1}) \right\rfloor \bmod N/2 \qquad (13)$$

and apply the opposite shift to normalize the curve:

$$n_p(\mathbf{z}) = S_{-p(\mathbf{z})}(\mathbf{z}) \qquad (14)$$

It is shown in [2] that the above normalization is invariant to starting point, except for an ambiguity in the standard circular shift, which may cause an extra shift of $N/2$:

$$p(\mathbf{z}') = (p(\mathbf{z}) + m) \bmod N/2 \qquad (15a)$$

$$p(n_p(\mathbf{z}')) = p(n_p(\mathbf{z})) = 0 \qquad (15b)$$

$$n_p(\mathbf{z}') = \begin{cases} n_p(\mathbf{z}), & 0 \le p(\mathbf{z}) + m < N/2 \\ S_{N/2}(n_p(\mathbf{z})), & N/2 \le p(\mathbf{z}) + m < N \end{cases} \qquad (15c)$$

The proposed selection of Fourier coefficients u_1 and u_{N-1} can detect *reflectional curve symmetries* and has also been employed in [7] for line pattern curves (e.g. character boundaries), while the use of other coefficients is also possible [6][9].

5. Rotation/Reflection Normalization

Rotation normalization is achieved by setting the phases of u_1 and u_{N-1} to zero, so that they become real and positive. In particular, assume that two curves $\mathbf{s}$, $\mathbf{s}'$ have been

orthogonalized and normalized w.r.t. starting point, thus satisfying (9). We then uniquely decompose matrix $\mathbf{Q}$ as

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \quad (16)$$

where $\theta \in [0, \pi)$, $s_x = \pm 1$, $s_y = \pm 1$, so that there is one-to-one relation between rotation/reflection parameters and elements of $\mathbf{Q}$. Adopting the complex vector notation $\mathbf{z}$, $\mathbf{z}'$,

$$\mathbf{z}' = (s_x \mathbf{x} + j s_y \mathbf{y}) e^{j\theta} \quad (17)$$

The rotation of curve $\mathbf{z}$ is normalized according to the average value of Fourier phases a_1 and a_{N-1}:

$$r(\mathbf{z}) = \left(\frac{1}{2}(a_1 + a_{N-1}) \right) \bmod \pi \quad (18a)$$

$$\mathbf{z}_1 = \mathbf{z}\, e^{-jr(\mathbf{z})} \quad (18b)$$

Then, horizontal and vertical reflection is normalized according to the 3rd-order moments of $\mathbf{z}_1$:

$$v(\mathbf{z}_1) = v_x(\mathbf{z}_1) + j v_y(\mathbf{z}_1) = \operatorname{sgn} m_{12}(\mathbf{z}_1) + j \operatorname{sgn} m_{21}(\mathbf{z}_1) \quad (19a)$$

$$n_r(\mathbf{z}) = \mathbf{z}_2 = v_x(\mathbf{z}_1)\mathbf{x}_1 + j v_y(\mathbf{z}_1)\mathbf{y}_1 \quad (19b)$$

where sgn denotes the signum function. It is then proved in [2] that $n_r(\mathbf{z})$ is invariant to rotation and reflection transformations:

$$r(\mathbf{z}') = (\lambda r(\mathbf{z}) + \theta) \bmod \pi \quad (20a)$$

$$n_r(\mathbf{z}') = n_r(\mathbf{z}) \quad (20b)$$

$$r(n_r(\mathbf{z}')) = r(n_r(\mathbf{z})) = 0 \quad (20c)$$

$$v_x(n_r(\mathbf{z}')) = v_y(n_r(\mathbf{z}')) = v_x(n_r(\mathbf{z})) = v_y(n_r(\mathbf{z})) = 1 \quad (20d)$$

where $\lambda = s_x s_y = \pm 1$. Note that, as in curve orthogonalization, the set of parameters $\{r(\mathbf{z}), v_x(\mathbf{z}), v_y(\mathbf{z})\}$ together with $n_r(\mathbf{z})$ contain all information about the original curve $\mathbf{z}$.

Two final normalization steps are required after rotation and reflection normalization. First, the starting point ambiguity of $N/2$ is resolved by applying an additional circular shift of $N/2$ samples if the starting point (x_0, y_0) lies to the left of the y-axis, i.e., $x_0 < 0$; otherwise the normalized curve is left intact. Second, the curve orientation is normalized to counterclockwise. Combining all the above results, it can be seen that curve $n_r(n_p(n_a(\mathbf{z})))$ obtained by the entire normalization procedure is *invariant to any affine transformation*.

6. Experimental Results

The performance of the proposed algorithm is evaluated using a shape database created from images and video sequences. Object contours have been obtained using the M-RSST color segmentation algorithm with the assistance of motion segmentation (in the case of video sequences), performed on the basis of 2-D parametric motion models. In the sequel, normalization results are given, for all proposed normalization steps and the algorithm's efficiency is discussed for affine transformations of (*i*) the same object under different, non-uniform sampling, (*ii*) similar objects and (*iii*) substantially different objects.

The first case is illustrated in Figure 1. A contour of a fish consisting of 100 sample points is depicted in Figure 1(a), along with two other curves obtained through arbitrary affine transformations under different, non-uniform sampling. The

normalization results are depicted in Figures 1(b,c,d). It can be seen that the final curves match very well, although normalization of each curve is performed *without knowledge* of the others. The slight discrepancies are due to different sampling; the match is perfect when the initial curves are affine transformations of the same sample curve.
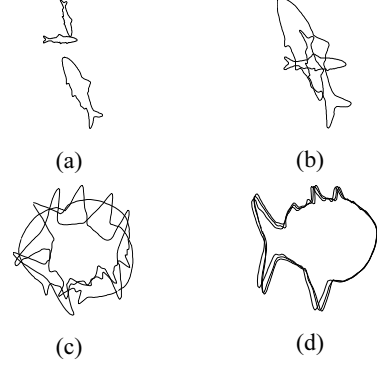

(a)   (b)

(c)   (d)

Figure 1. (a) Original fish contour and a pair of arbitrary affine transformations under different sampling, (b,c,d) curves after translation, skew/scaling and rotation normalization respectively.
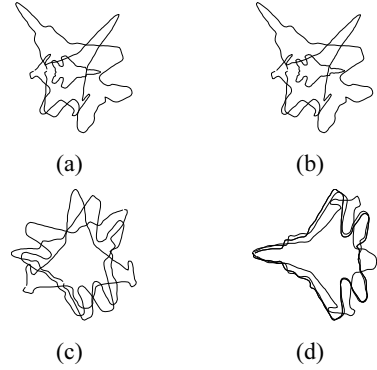

(a)   (b)

(c)   (d)

Figure 2. Normalization results for spatially similar object contours: (a) original contours for 3 distinct planes, (b,c,d) curves after translation, scaling and rotation normalization respectively.
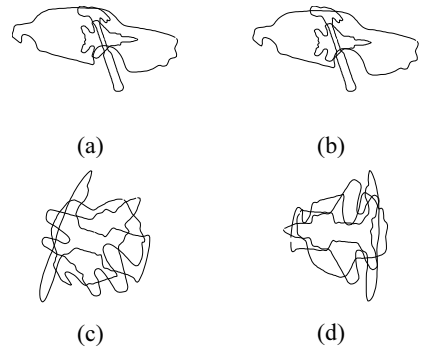

(a)   (b)

(c)   (d)

Figure 3. Contour alignment for substantially different objects: (a) original contours of a hammer, a car and a plane, (b,c,d) curves after translation, scaling and rotation normalization respectively.

Maybe the most important property of the proposed approach is its ability to align curves that appear to be spatially similar. The normalization results for three distinct airplanes are illustrated in Figure 2, where it can be seen that the proposed approach achieved successful alignment. These results should be directly compared to those of Figure 3, where three sample curves belonging to distinct object classes are employed. The final curves of Figure 2(d) yield larger resemblance on the basis of any matching scheme, compared to those of Figure 3(d). The proposed approach is proved to yield adequate results even in the presence of significant amount of noise in the employed curves, as shown in Figure 4.

(a)   (b)

Figure 4. Normalization results in the presence of noise: (a) initial and (b) final curves for spatially relative objects.

Contour 1	Contour 2	Points	FD	MFD
		0.01	0.02	0.01
		0.19	0.12	0.11
		0.75	0.41	0.57
		0.89	0.62	0.65
		0.76	0.25	0.32

Table 1. Indicative contour classification distances.
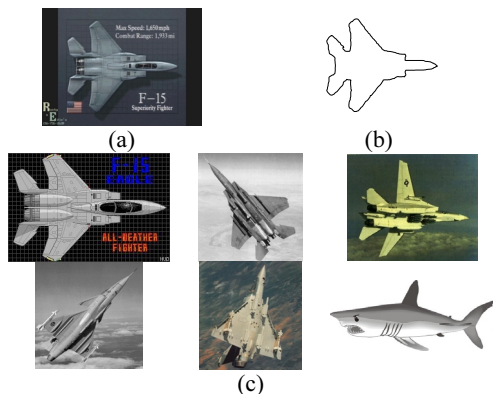

(a)   (b)

(c)

Figure 6. Image retrieval results (query-by-example) based on contour similarity: (a) input image, (b) extracted contour, (c) retrieved images in descending contour similarity.

Some indicative shape classification results are included in Table 1. The normalized counterparts of contours '1' and '2' were compared on the basis of three simple metrics, namely Euclidean distance of (*i*) the respective point sets, (*ii*) the estimated Fourier descriptors and (*iii*) the modified Fourier descriptors proposed in [8]. As it was intuitively expected, all three metrics prove to be indicative of the normalized curve resemblance. Finally, the proposed algorithm's performance was successfully tested for content-based retrieval purposes based on contour similarity on a small database containing 50 still images

of five visually distinct object classes; namely, airplanes, cars, fish, hammers, glasses. In Figure 6, retrieval results are illustrated for an input image containing an airplane. It must be pointed out that by utilizing only the object contour attribute, a fish could yield higher resemblance to an F-15 aircraft than a Stealth aircraft would. Other image attributes such as color and texture should thus be included in an integrated content-based retrieval system.

7. Conclusion

Using the curve normalization procedure presented in this paper, it is possible to obtain an affine-invariant curve representation without any actual loss of information. The procedure can be applied as a pre-processing step to any shape representation, classification or recognition technique, since it decouples affine-invariant description from feature extraction and pattern matching. This is verified by employing a number of curve similarity measures in the context of content-based retrieval from image and video databases. The proposed technique is experimentally shown to be considerably robust to noise and shape deformations. Moreover, since its computational cost is negligible, it can be integrated into any real-time system for content-based retrieval or even video coding.

8. References

[1] Y. Avrithis, A. Doulamis, N. Doulamis and S. Kollias, "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases," *Computer Vision and Image Understanding* **75** (1/2), pp. 3-24, 1999.

[2] Y. Avrithis, Y. Xirouhakis and S. Kollias, "Affine-Invariant Curve Normalization for Object Shape Representation, Classification and Retrieval," *Pattern Recognition Letters* (submitted for publication).

[3] A.D. Bimbo and P. Pala, "Visual Image Retrieval by Elastic Matching of User Sketches," *IEEE Trans. PAMI* **19** (2), pp. 121-132, 1997.

[4] F.S. Cohen, Z. Huang and Z. Yang, "Invariant Matching and Identification of Curves using B-Splines Curve Representation," *IEEE Trans. Image Processing* **4** (1), pp. 1-10, 1995.

[5] ISO/IEC JTC1/SC29/WG11, "MPEG-7: Context and Objectives (v.5)," Doc. N1920, 1997.

[6] F. Marques, B. Llorens and A. Gasull, "Prediction of Image Partitions Using Fourier Descriptors: Application to Segmentation-Based Coding Schemes," *IEEE Trans. PAMI* **7** (4), pp. 529-542, 1998.

[7] E. Persoon, and K.-S. Fu, "Shape Discrimination Using Fourier Descriptors," *IEEE Trans. PAMI* **8** (3), pp. 388-397, 1986.

[8] Y. Rui, A. She and T.S. Huang, "A Modified Fourier Descriptor for Shape Matching in MARS," *Image Databases and Multimedia Search. Series on Software Engineering and Knowledge Engineering* **8**, S.K. Chang (Ed.), World Scientific Publishing House in Singapore, pp. 165-180, 1998.

[9] D. Shen and H. Ip, "Generalized Affine Invariant Image Normalization," *IEEE Trans. PAMI* **19** (5), pp. 431-440, 1997.