Moment-Based Pattern Representation Using Shape and Grayscale Features*

Mikhail Lange, Sergey Ganebnykh, and Andrey Lange

Computing Centre RAS

Abstract. A moment-based approach is developed to constructing tree-structured descriptions of patterns given by region-based shapes with grayscale attributes. The proposed representation is approximately invariant with respect to the pattern rotation, translation, scale, and level of brightness. The tree-like structure of the pattern representations provides their independent encoding into prefix code words. Due to this fact, a pattern recognition procedure amounts to decoding a code word of the pattern by the nearest code word from a tree of the code words of selected templates. Efficient application of the pattern representation technique is illustrated by experimental results on signature and hand gesture recognition.

1 Introduction

Pattern representation is one of the basic problems in pattern recognition. In many cases, this problem is solved by constructing invariant descriptions of patterns with respect to their similarity transformations. In addition, it is necessary to construct the structured pattern descriptions that permit to decrease a computational complexity of a recognition procedure against a full search for a decision. Our goal consists in developing a technique of constructing the invariant tree-structured representations for a wide class of patterns given by 2D solid shapes with grayscale features.

A survey of known approaches to 2D shape representation is given in [5]. Among the developed shape descriptors a significant part refers to moment-based techniques [6] and the techniques based on shape decomposition into geometric primitives [2], [3]. A problem of appropriate fitting the objects by primitives of a given shape is considered in [9]. Of particular interest is a recursive decomposition approach providing tree-structured descriptions of shapes. The representations based on such approach are suggested in papers [1] and [4]. However, in many applications, the patterns are given by both shape and grayscale features. The examples are signatures, hand gestures, handwritten sings, and trade marks with nonuniform brightness. Thus, a demand arises for developing a pattern representation technique that combines the shape and grayscale features, and provides a similarity transformation invariance and a tree-like structure of

* This work is supported by the Russian Foundation for Basic Research, project 06-01-00524.

J. Martí et al. (Eds.): IbPRIA 2007, Part I, LNCS 4477, pp. 523–530, 2007.

the pattern description. This kind of pattern representations is convenient for their fast matching in pattern recognition. For these representations, the problem of pattern matching amounts to compare the representing trees by using a tree distance similar to metrics proposed in [7].

In this paper, we propose a technique of constructing tree-structured pattern representation on the basis of a recursive scheme of pattern partitioning and an approximation of the pattern parts (segments) by elliptic primitives using central and axial moments of inertia. The invariance of the representation is achieved by calculating the primitives in principal axis of the segments and by normalizing parameters of the primitives. A new dissimilarity measure is suggested in a space of the pattern representations. Also, an efficient application of the pattern representation technique is demonstrated by experimental results of signature and gesture recognition.

2 Statement of the Problem

Given grayscale image in the Cartesian coordinates X and Y, let a pattern be defined by a set of N pixels

$$P = \left\{ \, p_k = \{(x,y,z) : z(x,y) = z_k, |x - x_k| \le \frac{\Delta}{2}, |y - y_k| \le \frac{\Delta}{2}\}, k = \overline{1,N} \, \right\} \quad (1)$$

where $z(x,y)$ is a darkness function; Δ is a linear size of the pixel p_k; (x_k, y_k) are the coordinates of the center and $0 < z_k \le q$ is the darkness integer value of p_k. The zero darkness is assigned to background pixels. Let U and V be the Cartesian coordinates connected with the coordinates X and Y by a transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} c_{ux} & c_{uy} \\ c_{vx} & c_{vy} \end{pmatrix} \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix} \quad (2)$$

where (x^*, y^*) is a translation point and $\mathbf{c}_u = (c_{ux}, c_{uy})$ and $\mathbf{c}_v = (c_{vx}, c_{vy})$ are unit direction vectors of the coordinate axes U and V relative to the axes X and Y, respectively. In the coordinates U and V, we define an elliptic primitive by a set of points

$$Q = \left\{ (u, v, z^*) : \; z^*(u,v) = z^*, \; \frac{u^2}{r_u^2} + \frac{v^2}{r_v^2} \le 1 \right\} \quad (3)$$

where $z^*(u,v) = z^* > 0$ is a uniform darkness function and $r_u > 0$ and $r_v > 0$ are the radii along the appropriate axes U and V. According to (2) and (3), the primitive is determined by the following parameters

$$(x^*, y^*, z^*), (\mathbf{c}_u, \mathbf{c}_v), (r_u, r_v) \; . \quad (4)$$

An error of approximation of the object P by the primitive Q is defined by

$$E(P,Q) = \frac{1}{\|P \cup Q\|} \sum_{P \cup Q} \frac{|z - z^*|}{\max(z, z^*)}$$

where the sum is taken over all pixels which centers belong to the union $P \cup Q$ and $\|P \cup Q\|$ is the number of these pixels. It is assumed that z and z^* possess the zero values in all points outside P and Q, respectively. If $0 \leq \delta < 1$ is a given admissible approximation error, then the criterion of approximation is

$$E(P, Q) \leq \delta \ . \tag{5}$$

A scheme of the approximation of the total pattern P by a set of Q-primitives is based on partitioning the pattern into a set of segments and fitting the segments by the primitives satisfying (5). In what follows, we describe the technique of pattern representation (section 3), the scheme of pattern recognition based on the representation technique (section 4), and the results of experiments on signature and gesture recognition (section 5).

3 Recursive Moment-Based Pattern Representation

Due to recursiveness of the proposed pattern representation, it is enough to describe the moment-based approximation of any segment of the pattern and then to formalize the scheme of pattern partitioning. The approximation is based on calculating the extreme moments of inertia for both the object P (segment or total pattern) and the primitive Q given by (1) and (3), respectively. For the object P, the central moment of inertia relative to a point (x^*, y^*) is

$$J_c(P) = \sum_{k=1}^{N} \int_{p_k \in P} \left[(x - x^*)^2 + (y - y^*)^2 \right] w_z(x, y) \, dxdy \tag{6}$$

and the axial moment of inertia relative to an axis determined by the point (x^*, y^*) and the unit vector $\mathbf{c} = (c_x, c_y)$ is defined by

$$J_a(P) = \sum_{k=1}^{N} \int_{p_k \in P} \left[c_x(y - y^*) - c_y(x - x^*) \right]^2 w_z(x, y) \, dxdy \tag{7}$$

where $w_z(x, y) = z(x, y) \Big/ \sum_{k=1}^{N} \int_{p_k \in P} z(x, y) \, dxdy$ is the darkness distribution function for the object P.

Minimization of the moment (6) over the variables x^* and y^* yields the optimum point $(x^* = \sum_{k=1}^{N} x_k \tilde{w}_k, \ y^* = \sum_{k=1}^{N} y_k \tilde{w}_k)$ with the weights $\tilde{w}_k = z_k \Big/ \sum_{i=1}^{N} z_i$, $k = \overline{1, N}$. The minimum and maximum values of the moment (7) correspond to a pair of orthogonal axes (U, V) with the origin (x^*, y^*) and the unit direction vectors $\mathbf{c}_u = (c_{ux}, c_{uy})$ and $\mathbf{c}_v = (c_{vx}, c_{vy})$. These vectors are determined by a matrix of the second order central moments

$$\mathbf{G} = \begin{pmatrix} g_{yy} & -g_{xy} \\ -g_{yx} & g_{xx} \end{pmatrix}$$

with the elements $g_{xx} = \sum_{k=1}^{N} (x_k - x^*)^2 \tilde{w}_k + \frac{\Delta^2}{12}$, $g_{yy} = \sum_{k=1}^{N} (y_k - y^*)^2 \tilde{w}_k + \frac{\Delta^2}{12}$, and $g_{xy} = g_{yx} = \sum_{k=1}^{N} (x_k - x^*)(y_k - y^*) \tilde{w}_k$.

The eigenvalues are equal to $\lambda_{u,v} = \frac{1}{2}(g_{xx} + g_{yy}) \mp \frac{1}{2}\sqrt{(g_{xx} - g_{yy})^2 + 4g_{xy}g_{yx}}$ and real due to symmetry of the matrix $\mathbf{G}$. Since $\mathbf{G}$ is positive definite ($\det \mathbf{G} > 0$), therefore $\lambda_v \geq \lambda_u > 0$. If $\lambda_v > \lambda_u$, then the corresponding eigenvectors $\mathbf{c}_u$ and $\mathbf{c}_v$ determine the directions of the unique axes U and V. Notice that there are four different pairs $(\mathbf{c}_u, \mathbf{c}_v)$ for the given pair (λ_u, λ_v). We choose the decision $(\mathbf{c}_u, \mathbf{c}_v)$ corresponding to the minimum right rotation in the transformation (2). The found point (x^*, y^*) and eigenvectors $\mathbf{c}_u$ and $\mathbf{c}_v$ provide the following extreme moments of inertia

$$J_u(P) = \lambda_u, \qquad J_v(P) = \lambda_v, \quad J_c(P) = \lambda_u + \lambda_v \ . \tag{8}$$

For the primitive Q of the form (3), the moments of inertia relative to the found principal axes U and V are determined by the weighted mean values

$$J_u(Q) = \int_Q v^2 w_z^*(u,v)\, dudv = \frac{r_v^2}{4}, \quad J_v(Q) = \int_Q u^2 w_z^*(u,v)\, dudv = \frac{r_u^2}{4} \tag{9}$$

with the darkness density $w_z^*(u,v) = z^*(u,v) \Big/ \int_Q z^*(u,v)\, dudv$ that is equal to $1 \Big/ \int_Q dudv$ for the uniform function $z^*(u,v) = z^*$ of the primitive Q. The parameters (x^*, y^*) and $(\mathbf{c}_u, \mathbf{c}_v)$ of the primitive Q are determined by the principal axes U and V of the object P. The radii (r_u, r_v) follow from conditions of a moment-based equivalency of P and Q : $J_u(Q) = J_u(P)$; $J_v(Q) = J_v(P)$ and, using (8) and (9), these radii are equal to $r_u = 2\sqrt{\lambda_v}$ and $r_v = 2\sqrt{\lambda_u}$. The darkness value z^* of the primitive Q is found by the mean value $z^* = \sum_{k=1}^{N} z_k \,/N$ that yields a minimum mean square deviation of the darkness values for all pixels of the object P.

The pattern representation is constructed by the following recursive scheme. At the zero level ($l = 0$), the total pattern P is regarded as the object P_0 with the number $n = 0$. The object P_0 is approximated by the finest matched primitive Q_0 as described above. Given admissible error δ, if the pair (P_0, Q_0) satisfies the criterion (5) or P_0 consists of a single pixel, the primitive Q_0 is marked as "end" node. Otherwise, the object P_0 is partitioned into two segments by the principal axis V that moment of inertia $\lambda_v > \lambda_u$. The obtained segments are regarded as the new objects P_1 and P_2 of the first level $l = 1$. The described procedure is repeated for the objects P_1 and P_2 and for the new objects of the next levels. In general case, the object P_n of the l-th level produces two new objects P_{2n+1} and P_{2n+2} of the $(l+1)$-th level. The maximum level of partitioning is upper bounded by a given L and all primitives of the L-th level are marked as "end nodes". Note that if $\lambda_u = \lambda_v$ for some object, then the pair (λ_u, λ_v) does not give the unique pair of eigenvectors $(\mathbf{c}_u, \mathbf{c}_v)$. For this object, the pair of axes (U, V) is assigned by the principal axes of the total pattern. The last notice limits a class of admissible patterns which matrices $\mathbf{G}$ have different eigenvalues.

Given pattern P, the described recursive scheme produces the pattern representation R in a form of a complete binary tree of the primitives (nodes)

$$R = \{Q_n : 0 \leq n \leq n_{\max}\} \tag{10}$$

where n is the node number of the level $l_n = \lfloor \log_2(n+1) \rfloor \leq L$. Each node Q_n in the tree (10) is described by the parameters (4) that are recalculated into the principal coordinates $(U = U_0, V = V_0)$ of the root node Q_0 by using the transformation (2). The recalculated centers (u^*, v^*), radii (r_u, r_v), and darkness values z^* of the primitives are normalized for providing scale and brightness invariance of the representation (10).

Two examples of the representations are shown in Fig. 1 for signature and hand gesture. The left pictures correspond to real images, the middle pictures show the extracted patterns ($q = 255$ for signature and $q = 120$ for gesture), and the right pictures illustrate the pattern representations by the grayscale elliptic primitives ($\delta = 0.05$, $L = 7$). Increasing L provides more detailed representation.

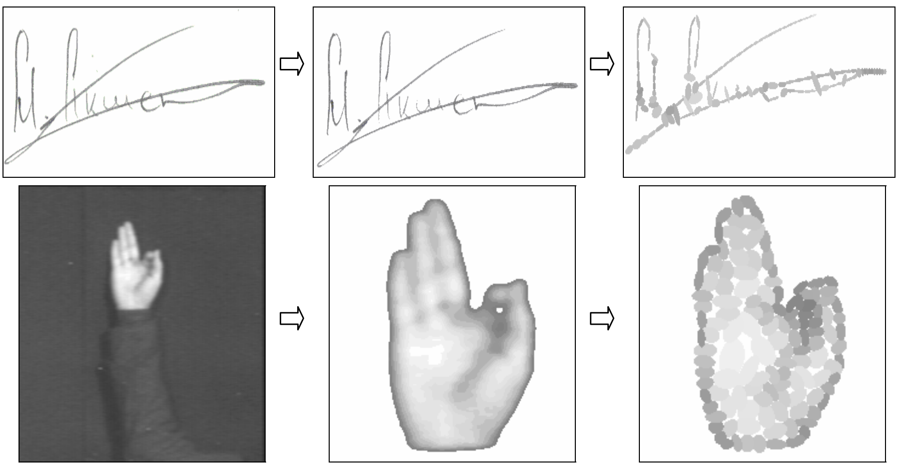

Fig. 1. Examples of the representations for grayscale signature and hand gesture

4 Application for Pattern Recognition

Pattern recognition by the nearest template requires a dissimilarity measure for any pair of the pattern representations. This measure is based on defining both a correspondence between the nodes of a pair of trees and an intersection of the trees. For the trees $(R, \hat{R})$ of the form (10), the nodes $Q_n \in R$ and $\hat{Q}_n \in \hat{R}$ are regarded corresponding to each other if these nodes have the same numbers. A set of the corresponding nodes gives the intersection $R \cap \hat{R}$.

Let $\rho(Q_n, \hat{Q}_n) \geq 0$ be a dissimilarity function of the corresponding nodes Q_n and $\hat{Q}_n$ for the pairs $(Q_n, \hat{Q}_n) \in (R \cap \hat{R})$. Using this function, we define a loss function

$$d(Q_n, \hat{Q}_n) = \begin{cases} \rho(Q_n, \hat{Q}_n), & \text{if } Q_n \text{ and/or } \hat{Q}_n \text{ are "end" nodes,} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Then, the dissimilarity measure of the trees $(R, \hat{R})$ is defined by

$$D(R, \hat{R}) = \sum\nolimits_{R \cap \hat{R}} d(Q_n, \hat{Q}_n) w(Q_n, \hat{Q}_n) \tag{12}$$

where $w(Q_n, \hat{Q}_n) = 2^{-l_n}$ and the sum is taken over all pairs $(Q_n, \hat{Q}_n) \in (R \cap \hat{R})$. The measure (12) requires a definition of the function $\rho(Q_n, \hat{Q}_n)$ for the ellipses Q_n and $\hat{Q}_n$ given in the same principal coordinate axes U and V of the root nodes Q_0 and $\hat{Q}_0$. A pair of recalculated and normalized primitives Q_n : $((u_n^*, v_n^*, z_n^*), (\mathbf{c}_{nu}, \mathbf{c}_{nv}), (r_{nu}, r_{nv}))$ and $\hat{Q}_n$: $((\hat{u}_n^*, \hat{v}_n^*, \hat{z}_n^*), (\hat{\mathbf{c}}_{nu}, \hat{\mathbf{c}}_{nv}), (\hat{r}_{nu}, \hat{r}_{nv}))$ is shown in Fig. 2. According to Fig. 2, the ellipses Q_n and $\hat{Q}_n$ are determined

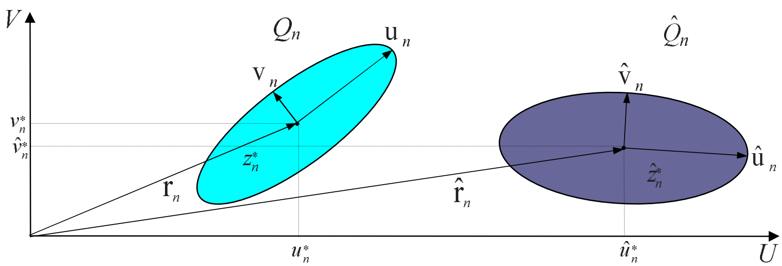

Fig. 2. A pair of primitives Q_n and $\hat{Q}_n$ in principal axes U and V of the pattern

by the center vectors $\mathbf{r}_n = (u_n^*, v_n^*)$ and $\hat{\mathbf{r}}_n = (\hat{u}_n^*, \hat{v}_n^*)$, and the appropriate pairs of the direction vectors $(\mathbf{u}_n = r_{nu}\mathbf{c}_{nu}, \mathbf{v}_n = r_{nv}\mathbf{c}_{nv})$ and $(\hat{\mathbf{u}}_n = \hat{r}_{nu}\hat{\mathbf{c}}_{nu}, \hat{\mathbf{v}}_n = \hat{r}_{nv}\hat{\mathbf{c}}_{nv})$, respectively. Taking into account the above vectors, we choose

$$\rho^2(Q_n, \hat{Q}_n) = \frac{|\mathbf{r}_n - \hat{\mathbf{r}}_n|^2}{\max^2(|\mathbf{r}_n|, |\hat{\mathbf{r}}_n|)} + \frac{|\mathbf{u}_n - \hat{\mathbf{u}}_n|^2 + |\mathbf{v}_n - \hat{\mathbf{v}}_n|^2}{\max(|\mathbf{u}_n|^2 + |\mathbf{v}_n|^2, |\hat{\mathbf{u}}_n|^2 + |\hat{\mathbf{v}}_n|^2)} + \frac{(z_n^* - \hat{z}_n^*)^2}{\max^2(z_n^*, \hat{z}_n^*)}$$

that coupled with (11) completely defines the measure (12).

Selection of the templates is performed at the stage of training the classifier. For this goal, we use a training set $\mathbf{R}_0$ that contains a fixed number of semantic groups of patterns with a given number of the pattern representations in each group. Processing of the semantic groups in $\mathbf{R}_0$ is performed independently and produces a fixed number m of the selected representations in each group. In the result, a set of the templates $\mathbf{R}^m \subset \mathbf{R}_0$ is constructed by combining the templates selected in each semantic group. In case of $m = 1$, the single template $\hat{R}$ in the given group of $\mathbf{R}_0$ yields $\sum_i D(R_i, \hat{R}) = \min_j \sum_i D(R_i, R_j)$, where the sums are taken over all pattern representations of the given semantic group. The template $\hat{R}$ provides the smallest dissipation $\sigma = \sum_i D(R_i, \hat{R}) / \sum_i 1$ per one sample of the group in $\mathbf{R}_0$. In case of $m > 1$, the selected templates $\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_m$ in each semantic group of $\mathbf{R}_0$ yield the conditional maximum $\sum_{i=1}^m \sum_{j=i+1}^m D(\hat{R}_i, \hat{R}_j) = \max \sum_{i=1}^m \sum_{j=i+1}^m D(R_i, R_j)$ over all possible sequences $R_1, R_2, \ldots, R_m$ satisfying the inequality $\sum_{i=1}^m D(R_i, \hat{R}) \le m\sigma$.

Given $\mathbf{R}^m$, the recognition decision about a pattern P via the pattern representation R is made by the nearest template $\hat{R}^* \in \mathbf{R}^m$ providing $D(R, \hat{R}^*) = \min_{\hat{R} \in \mathbf{R}^m} D(R, \hat{R})$. The recognition efficiency is determined by a probability $P_{\text{true}} = \Pr\{\text{"true decision"}\}$ that is a ratio of a number of true decisions to a total number of patterns being under recognition.

Any representing tree R of the form (10) can be written as a code word which length of the code word is proportional to the number of the nodes in R. Due to completeness of the representing trees, a given finite set of the code words satisfies the Kraft inequality [8]. This property yields the set $\mathbf{R}^m$ in a form of a tree $T(\mathbf{R}^m)$, in which the search for the nearest template can be performed by a modification of the Viterbi sequential decoding algorithm [8]. This step by step algorithm processes a limited number of subpaths in $T(\mathbf{R}^m)$ at each step. If the length of the subpaths is equal to $K = \gamma \log N$, where $\gamma > 0$ and $N = 2^{L+1}$ is the upper estimation of the code word length, then the computational complexity of the search algorithm is upper bounded by $O(N^{\gamma+1})$.

5 Experimental Results

The utility of the developed pattern representation technique was confirmed by experiments on recognition of signatures and hand gestures given by real grayscale images. The signatures were kindly submitted by our students and the gestures were the letters of American Sign Language (ASL) taken from the website http://www.vision.auc.dk/~tbm/Gestures/database.html.

The training set $\mathbf{R}_0$ of the signatures consisted of 32 (number of persons) semantic groups with 4 signatures of one person in each group. The test set of signatures submitted for recognition contained the same 32 semantic groups with 4 other samples of signatures in each group. In the case of hand gestures, the number of semantic groups was determined by the size of the downloaded ASL alphabet and it was equal to 25 signs. Each semantic group in the gesture training set $\mathbf{R}_0$ was given by 8 samples and the test set of gestures contained the same 25 signs per 8 other samples in each group. In both cases, the sets of the templates $\mathbf{R}^2 \subset \mathbf{R}_0$ $(m = 2)$ were used. For $\delta = 0.05$ and different values L, the results in terms of P_{true} are given by table in Fig. 3. As shown in the table, the pattern representations based on shape and grayscale features $(q > 1)$ provide a profit relative to the representations using only shape features $(q = 1)$. As a whole, the recognition efficiency grows when L increases and, for fixed L, it can be improved by increasing the parameter m.

δ=0.05	Gestures		Signatures	
	grayscale (q=120)	binary (q=1)	grayscale (q=255)	binary (q=1)
L=6	0.980	0.975	0.977	0.969
L=7	0.990	0.980	0.992	0.969
L=8	0,995	0.980	0.992	0.977
L=9	0,995	0.985	0.992	0.992

Fig. 3. Experimental results of gesture and signature recognition

6 Conclusion

In this paper, we proposed the new technique for constructing the tree-structured representations of region-based grayscale patterns. The basis of the technique consists of the recursive pattern decomposition and moment-based approximation of the pattern segments by elliptic primitives taken in principal axis of the segments. The last property provides the rotation and translation invariance of the pattern representation. The scale and brightness invariance is achieved by appropriate normalization of the primitives. As compared with classical descriptors using Zernike moments or Fourier expansions, our technique is based on zero, first, and second geometric moments and it is meant for structured recognition algorithms. The tree-like structure of the pattern representations permits to construct any set of templates as a tree of code words and to search for the decision templates by a scheme of sequential decoding. The search algorithm has a polynomial computational complexity of the length of the code words. Moreover, a multiresolution property of the proposed representation gives a chance to accelerate the recognition procedure using a scheme of successive refinement. The experiments on recognition of real signatures and ASL gestures showed the probability of true decisions within $0.98 - 0.99$. We plan to develop our technique with other primitives and measures as well as to make experiments on handwritten sign recognition. Also, an accelerated recognition algorithm based on the multiresolution property of the pattern representations will be researched.

References

1. Berretti, S., Del Bimbo, A.: Multiresolution spatial partitioning for shape representation. IEEE Proceedings of ICPR 2, 775–778 (2004)
2. Jagadish, H.V., Bruckstein, A.M.: On sequential shape descriptions. Pattern Recognition 25, 165–172 (1992)
3. Kim, H., Park, K., Kim, M.: Shape decomposition by collinearity. Pattern Recognition Letters 6, 335–340 (1987)
4. Lange, M.M., Ganebnykh, S.N.: Tree-like Data Structures for Effective Recognition of 2-D Solids. IEEE Proceedings of ICPR 1, 592–595 (2004)
5. Loncaric, S.: A survey of shape analysis techniques. Pattern Recognition 34(8), 983–1001 (1998)
6. Prokop, R.J., Reeves, A.P.: A survey of moment-based techniques for unoccluded object representation and recognition. CVGIP: Graphical Models and Image Processing 54, 438–460 (1992)
7. Torsello, A., Hodovic, D., Pelillo, M.: Four metrics for efficiently comparing attributed trees. IEEE Proceedings of ICPR 2, 467–470 (2004)
8. Viterbi, A.J., Omura, J.K.: Principles of Digital Communication and Coding. McGraw-Hill, New York (1979)
9. Voss, K., Suesse, H.: Invariant fitting of planar objects by primitives. IEEE Proceedings of ICPR, pp. 508–512 (1996)