2011 International Conference on Document Analysis and Recognition

# Symbol Recognition By Multiresolution Shape Context Matching

Feng Su, Tong Lu, and Ruoyu Yang
*State Key Laboratory for Novel Software Technology*
*Nanjing University*
*Nanjing, 210093, China*
*suf@nju.edu.cn*

***Abstract*—We present a multiresolution scheme for symbol representation and recognition based on statistical shape features. We define a symbol as a set of shape points, each of which is then described by a pyramid of shape context features. The pyramid is constructed by successively partitioning the image surrounding one shape point into increasingly finer sub-regions and computing the local shape context descriptor inside each sub-region. To recognize a symbol, we compute the optimal matching between symbol prototypes and the image region, based on the weighted distance measurements across various scales. We also define an adaptive surround suppression measure that assigns different weights to the shape point depending on the complexity of its surrounding context, so as to reduce the effect of local intersections to shape matching. The experimental results show the effectiveness of the proposed shape context pyramid matching method as well as its promising aspects in handling intersecting symbols.**

***Keywords*-symbol recognition; shape context; multiresolution; surround suppression;**

## I. Introduction

Graphic symbols, i.e. symbols made of lines, arcs and simple geometric primitives, can be found in many types of graphic drawings. Usually symbols play an important role in interpreting and making effective use of the drawings, thereby should be recognized from the image with high precision and efficiency.

Many symbol recognition methods have been proposed in the past two decades [1]–[4], which can be essentially sorted into two groups: statistical approaches and structural approaches. Compared to the latter, statistical approaches employ pixel-level descriptors, which are robust to local noises and require no detection of symbol components in advance, thus are widely exploited. However, statistical descriptors are usually vulnerable to structural changes, such as intersecting between objects that are common in real graphical documents. In such cases, it's possible for the local parts of the object to preserve their shape integrity, which gives justification to the multiresolution methods.

Multiresolution analysis has been an effective processing method to vision problems and is widely exploited. In shape-based object modeling, Del Bimbo and Pala [5] present a multi-scale hierarchical shape representation in which shape details are progressively filtered out while shape characterizing elements are preserved. Grauman and Darrell [6] propose a pyramid matching method to find an approximate correspondence between two sets of vectors in a $d$-dimensional feature space. The multiresolution histogram places a sequence of increasingly coarser grids over the feature space, so that varies the resolution at which the features are computed, but the histogram resolution stays fixed. In [7], contrarily, the former is fixed while the spatial resolution is varied.

In this paper, we propose a spatial multiresolution scheme for describing and recognizing graphic symbols based on their shapes. We partition the image surrounding the shape point into increasingly finer sub-regions and computing histograms of local features found inside each sub-region. The resulting is a pyramid of feature histograms, which is used in robust matching of symbol shape points. We also define an adaptive weighting measure based on the surround suppression effect for every point extracted from the target image region, to reduce the effect of local intersections to shape matching.

The paper is organized as follows. In Section II, we give the formulation of the multiresolution symbol representation including the shape feature employed and the construction of the histogram pyramid. In Section III, we describe details of the pyramid-based shape matching as well as the surround suppression measure. In Section IV, we present some experiment results of the proposed method and possible improvements.

## II. Multiresolution Symbol Representation

Multiscale processing is an old but powerful idea, which is usually applicable whenever one wishes to implement an algorithm that involves iterative coarse-to-fine processing or the object of interest exhibits different properties when analyzed at different resolutions.

For linear symbols addressed in our work, we extract their contour or skeleton point set and sample them uniformly for the reduced data size. Then, we extract shape features, the shape context [8] in this work, at different ROI scales, resulting in a higher-dimensional multiresolution representation of the symbol that preserves more information.

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.265

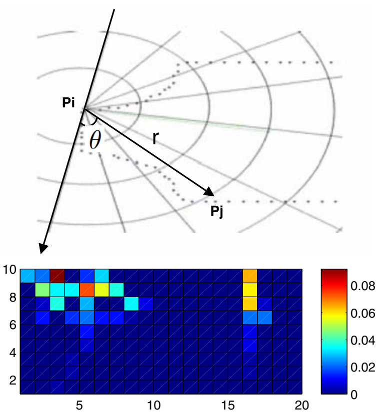


Figure 1. Shape contexts: the log-polar coordinates used in computing the shape contexts (upper) and an example shape context histogram (lower).

### A. Shape Contexts

Shape Context [8] proposed by Belongie is a rich local descriptor of object shape, which is represented as an unordered discrete set of $N$ points $\mathcal{P} = \{p_1, \ldots, p_N\}, p_i \in \Re^2$ samp1ed from the interna1 or externa1 object contours. For each point $p_i$ on the shape, considering the set of $N-1$ vectors that originate at $p_i$ and extend to $p_j$ $(j \neq i)$, a shape context $h_i$ depicts the distribution of the positions of a11 $p_j$ relative to $p_i$ and is used as a compact description of the local object shape around $p_i$. In the experiments of Belongie et al., $h_i$ is formulated as a coarse 2-D histogram $h_i(k,l)$ of the radial length $r$ and angle $\theta$ of the vector $\overline{p_i p_j}$ in log-polar coordinates, where $k \in [1 \ldots Q_r]$, $l \in [1 \ldots Q_\theta]$ and $Q_r, Q_\theta$ are the number of quantization levels of the length $r$ and angle $\theta$ in histogram $h_i$ respectively. The illustration and example of the shape context are shown in Fig. 1.

The shape context is invariant to trans1ation by definition since a11 measurements are taken with respect to points on the object. It's also invariant to scale transformation by norma1izing the radial length $r$ based on the holistic size of the shape. For rotation invariance, the tangent direction at point $p_i$ can be used as the positive x-axis of the re1ative log-polar coordinate system.

With shape contexts, the matching of a point $p_i$ on the first shape with a point $q_j$ on the second shape is based on comparing their shape context histograms $h_i$ and $h_j$, using certain cost metric $C_{i,j} = \mathcal{C}(h_i, h_j)$ like the $L1$-norm (1) as the similarity measurement.

$$\mathcal{C}(h_i, h_j) = \sum_{k=1}^{Q_r} \sum_{l=1}^{Q_\theta} |h_i(k,l) - h_j(k,l)| \qquad (1)$$

For the two shapes composed of $M$ points $\{p_i\}_{i \in [1 \ldots M]}$ and $N$ points $\{q_j\}_{j \in [1 \ldots N]}$ respectively, given the set of costs $\{C_{i,j}\}$ between a11 pairs of points $(p_i, q_j)$, we compute the tota1 cost of matching as:

$$\overline{C}(\pi) = \sum_i C_{i,\pi(i)} \qquad (2)$$

where, $\pi$ is a permutation of points that minimizes (2). The optimal $\pi$ can be solved by methods for the weighted bipartite matching prob1em like the Hungarian algorithm. If the point sets have unequal cardinality ($M \neq N$), the points of the smaller set are mapped to some subset of the points in the larger set.

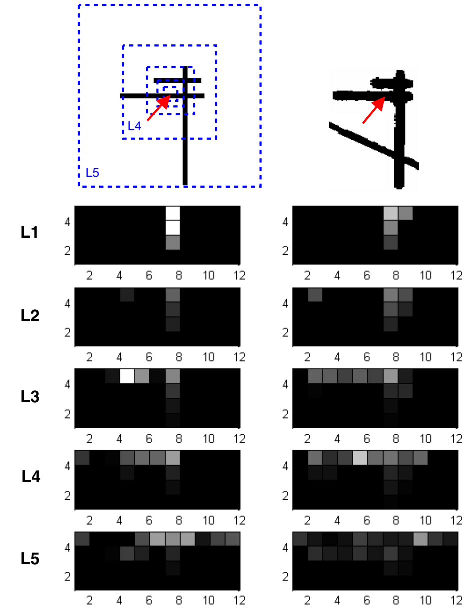


Figure 2. Illustration of multiresolution shape context pyramid. Left: one symbol prototype and the 5 shape context levels extracted from the finest region (L1) to the coarsest region (L5) surrounding the point at the arrowhead. Right: one corresponding symbol instance intersecting with another line and the shape context levels.

### B. The Shape Context Pyramid

Instead of extracting and analyzing shape features at a single resolution, we propose to construct a feature pyramid based on shape contexts to provide a robust representation of the symbol shape. Unlike the usual image pyramid methods employing a successively reduced version of the image at different detail levels, we extract the shape features at the single original image resolution, however, from a hierarchy of nested regions of interest of different sizes, as illustrated in Fig. 2.

Supposing we construct a $L$-level feature pyramid:

$$H_{p_i} = [h_i^1, h_i^2, \ldots, h_i^L]$$

for each shape point $p_i$. $h_i^L$ is the shape context computed at the coarsest level (the whole object) as described in the previous section, while $h_i^1$ is the finest level. Specifically, for each pyramid level $s \in [1 \ldots L]$, we compute the shape context $h_i^s$ based on the point subset $P_i^s = \{p_j | p_j \in \Re_s^2(p_i)\}$, where $\Re_s^2(p_i)$ is a rectangular window centered at $p_i$ and of the size $[2D_h^s, 2D_w^s]$:

$$D_h^s = \frac{D_H}{2^{L-s}} \qquad D_w^s = \frac{D_W}{2^{L-s}}$$

where, $D_H$ and $D_W$ are the height and width of the object, respectively.

The basic idea of computing a spatial pyramid of local shape features for every object point is to allow partial matching between symbol prototype and the instances intersecting with other graphical objects. For such symbols, the matching can be weak when measured at higher pyramid levels with a large ROI, which may involve much of the interfering part. At the lower level (smaller ROI), however, it's likely for unaffected parts to hold high matching accuracy, as shown by the similar distributions of shape context between two symbols at the finer pyramid levels in Fig. 2.

## III. Symbol Recognition By Shape Context Pyramid Matching

Based on the shape context pyramid model of the symbol, in this part, we consider the problem of matching a symbol prototype $T$ with the target image region $X$ that is roughly segmented from the input image. Our matching algorithm adopts the following steps:

1) Compute the shape context pyramid distance matrix $C_{i,j}$ ($i = 1..M, j = 1..N$) between $X$ and $T$ that have $M$ and $N$ shape points respectively, using the pyramid extension of Equation (1).
2) Look for the optimal matching or correspondence between the shape point sets of $X$ and $T$, which minimizes the total matching cost.
3) From a set of symbol prototypes $\{T_i\}$, determine the one with the minimum matching cost with $X$.

### A. Computation of Shape Context Pyramid Distance Matrix

Similar to (1), the pyramid distance matrix $C_{i,j}$ between two shapes defines the cost to match every pair of points $(p_i, q_j)$. Since each point is described by a histogram pyramid, the matching cost of two points $p_i$ and $q_j$ can have various forms. A basic form is the matching between corresponding levels:

$$C_{i,j} = \sum_{s=1}^{L} w_s \times \mathcal{C}(h_i^s, h_j^s) \tag{3}$$

where, $h_i^s$ and $h_j^s$ is the $s^{th}$-level shape context histogram of point $p_i$ and $q_j$ respectively, $\mathcal{C}()$ is defined by (1).

In (3), we associate a *scale weight*:

$$w_s = 1/2^{(L-s+1)} \tag{4}$$

with each pyramid level $s$, which increases with the level and is used in the matching measurement. Generally, features extracted at higher level, i.e. from larger ROI, are more robust and descriptive than those from smaller ROI, thereby we highlight the matching costs in coarse scales with a relatively larger weight.

Taking the matching between different pyramid levels into consideration, we can formulate it as a bipartite matching prob1em: the levels of $H_{p_i}$ and $H_{q_j}$ correspond to the two disjoint vertex sets of the bipartite graph, and the distance between histogram $h_i^s$ and $h_j^t$ of any two levels ($s$ and $t$) plays as the weighted edge $E_{s,t} = \mathcal{C}(h_i^s, h_j^t)$, so that $E_{s,t}$ being the inter-level cost matrix. To favor matching between close levels, an extra inter-level weight matrix $O_{s,t} = 1/2^{L-|s-t|}$ can be defined and the revised inter-level cost matrix is computed as:

$$C_{s,t} = E_{s,t} \cdot O_{s,t}$$

which can be solved as (2) to obtain an optimal permutation $\pi_l(s)$ of levels. Based on it, the point matching cost $C_{i,j}$ is computed by a permutation version of (3):

$$C_{i,j} = \sum_{s=1}^{L} (w_s w_{\pi_l(s)})^{\frac{1}{2}} \times \mathcal{C}(h_i^s, h_j^{\pi_l(s)}) \tag{5}$$

### B. Adaptive Point Weighting by Surround Suppression

The surround suppression method is originally proposed in [9] for contour detection in images. It is based on the results from neurophysiology which show that the existence of a complex surround decreases the perceptual importance of the point under concern in human visual system, for example, resulting in the decreased saliency of a contour in presence of surrounding texture.

Inspired by [9], we can introduce a *surround suppression weight* $wss_p$ in shape matching for every shape point $p$ in the target image region such that, for a point with strong surround suppression, like those reside in the intersecting area of objects, it makes a less significant contribution to the shape matching score, while for a point located in a clear neighbourhood, and thus more likely to be of the unaffected parts of object, a larger weight should be assigned to it.

Considering the shape context of a point has encoded the distribution of its neighbouring pixels, intuitively we can measure the strength of surround suppression and define the weight $wss_p$ based on it. The basic idea is to assign larger weights to those points that exhibit consistency and simplicity of surrounding context throughout various scales. Given the shape context pyramid $H_p = [h_p^1, \ldots, h_p^L]$ of a point $p$ in the target region, we define its weight $wss_p$ as the sum of two terms, weighted by $\beta$:

$$wss_p = \beta wss_p^e + (1-\beta) wss_p^h \tag{6}$$

The term $wss_p^e$ is the sum of the entropy of shape context distribution at each pyramid level of point $p$, which measures its intra level shape context variance:

$$wss_p^e = \sum_{s=1}^{L} w_s \frac{Entropy(\{h_p^s\})}{Ent_{max}}$$

where, $Entropy(\{h_p^s\}) = -\sum_{k,l} h_p^s(k,l)\log[h_p^s(k,l)]$, $Ent_{max} = \log(Q_r Q_\theta)$ is the maximum possible entropy of $h_p^s$.

The term $wss_p^h$ measures the inter level shape context variances of point $p$ and is computed as follows:

1) Compute the coherence $I_p^{s=1..L-1}$ of shape context distribution between adjacent pyramid levels $h_p^s$ and $h_p^{s+1}$:

$$I_p^s = \mathcal{I}(h_p^s, h_p^{s+1})$$

where, $\mathcal{I}()$ is the histogram intersection function measuring the overlap between two histograms' bins:

$$\mathcal{I}(A,B) = \sum_{j=1}^{R} \min(A(j), B(j)) \tag{7}$$

where $A$ and $B$ are histograms with $R$ bins, and $A(j)$ denotes the count of the $j^{th}$ bin of $A$.

2) Compute $wss_p^h$ as a weighted sum of $I_p^{s=1..L-1}$:

$$wss_p^h = \sum_{s=1}^{L-1} w_s(1.0 - I_p^s)$$

With the surround suppression weights, the point matching cost $C_{i,j}$ can be extended as:

$$C_{i,j} = wss_i \times C'_{i,j} \tag{8}$$

where, $wss_i$ is the surround suppression weight of the $i^{th}$ point in the target image region, $C'_{i,j}$ is computed using (3) or (5).

### C. Match Symbol Model with Image

In this part we match a candidate region $X$ in the input image with $K$ predefined symbol prototypes $T^{(k=1..K)}$, all represented as a set of shape context pyramids associated with the shape points. We compute the shape distance $\mathcal{D}(X, T^{(k)})$ by their total shape context pyramid distance $\overline{C}(X, T^{(k)})$ (2) based on the point-to-point matching costs $C_{i,j}$ described before, where $i$ refers to the $i^{th}$ point of the candidate region $X$ and $j$ refers to the $j^{th}$ point of one symbol prototype $T^{(k)}$.

The final symbol class of $X$ is determined by:

$$\arg\min_k \mathcal{D}(X, T^{(k)}) \tag{9}$$

For computational efficiency, the method proposed in [10] can be employed in place of the Hungarian algorithm used in this experimental work for approximate optimal point matching.

For more robustness of matching, as proposed in [8], we can also define the shape distance $\mathcal{D}(X, T^{(k)})$ as the sum of two terms, weighted by $\lambda$:

$$\mathcal{D}(X, T^{(k)}) = \lambda\overline{C}(X, T^{(k)}) + (1-\lambda)\overline{E}(X, T^{(k)}) \tag{10}$$

where, $\overline{C}(X, T^{(k)})$ is the shape context pyramid distance, $\overline{E}(X, T^{(k)})$ is the bending energy of the matching points of two shapes based on the thin plate spline (TPS) model.

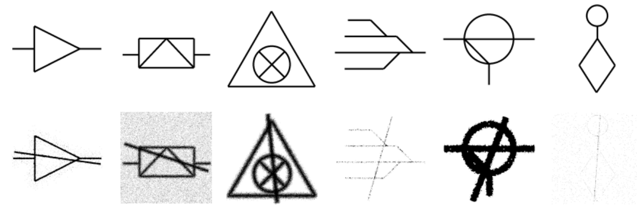

Figure 3. Synthetic intersecting symbol samples. Top: symbol prototypes. Bottom: synthetic symbol samples with randomly added intersecting lines in various degradation models.

## IV. RESULTS

In this section we tested the proposed shape context pyramid matching scheme in simulations on technical line symbols. The data sets from the symbol recognition contest of GREC2005 [11] and synthetic intersecting samples are used. Since the shape context is invariant to trans1ating, scaling and rotation, we focus on the first category of tests in [11] that includes 6 different models of noises and degradations to the symbol shape. To simulate intersecting symbols, we generate synthetic symbol samples by adding lines with random orientations and in the same degradation patterns to GREC2005 test images. Some of them are illustrated in Fig. 3.

For every symbol prototype and test image, we extract the skeleton points and uniformly subsample them to obtain the reduced shape point sets. Then, we compute the shape context pyramid using the same setting of shape context (5x12, 5 distance bins and 12 orientation bins) as in [8]. Three levels (high, moderate and low) of shape point subsampling resolutions are tested in the experiments, for which the average numbers of sample points per symbol are around 100, 70, 50, respectively. It's worth notice that, for most of the degradation models, the sample point sets include quite a number of noises and skeletonization artifacts, which could be eliminated or reduced for further increase of efficiency if appropriate preprocessings are taken.

Table I shows the average recognition accuracy of the proposed *pyramid matching* method on GREC2005 category-1-150 test samples with moderate subsampling resolution. Table II shows the accuracy of pyramid matching with surround suppression on synthetic intersecting test samples with $\beta = 0.5$ in (6). For simplicity, we use (3) for pyramid distance and set $\lambda = 1.0$ in (10), i.e., without using the bending energy in shape matching. We also present corresponding results using the plain shape context feature (without TPS-based deformable matching), which is extracted from the symbol holistically, for comparison.

Fig. 4 and Fig. 5 show the comparisons of the proposed and the shape context-based method on the average recognition accuracy over subsampling resolutions and degradation models, respectively.

The result shows that the proposed multiresolution anal-

Table I
AVERAGE ACCURACY ON GREC2005 TEST SETS WITH DISTORTION AND DEGRADATION (%).

| | Degradation Models | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| pyramid matching | 98 | 93 | 86 | 89 | 72 | 65 |
| shape context | 94 | 83 | 74 | 79 | 66 | 57 |

Table II
AVERAGE ACCURACY ON SYNTHETIC TEST SETS WITH INTERSECTING LINES (%).

| | Degradation Models | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| pyramid matching | 81 | 75 | 67 | 72 | 52 | 27 |
| shape context | 71 | 72 | 60 | 67 | 46 | 26 |

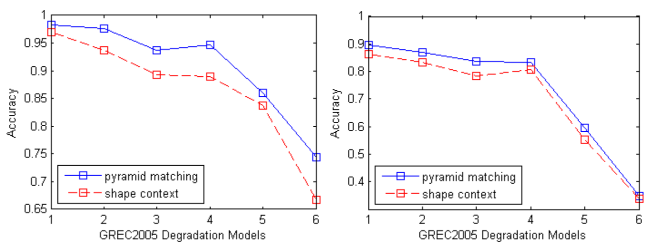


Figure 4. Average accuracy in each degradation model over all subsampling resolutions on GREC2005 (left) and synthetic (right) test sets.

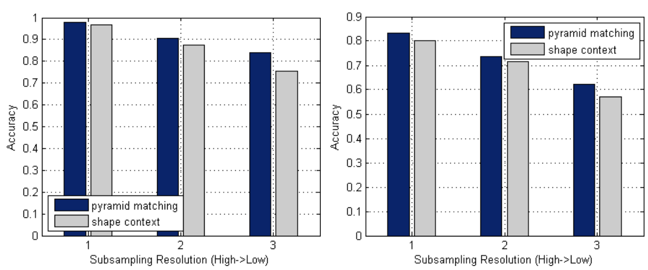


Figure 5. Average accuracy at variant subsampling resolution over all degradation models on GREC2005 (left) and synthetic (right) test sets.

ysis of symbol shapes improves the robustness of shape matching in presence of partial intersections and various degradations. It also shows the pyramid's smoothing effect on the noises and local deformations that could otherwise cause false matching of shape points. Like the shape context feature, the accuracy of the proposed method decreases with the reduction of shape points that results from coarser subsampling, though retains higher than the former on the majority of test sets inspected.

## V. CONCLUSION

We present a new spatial multiresolution scheme for describing and recognizing graphic symbols based on a pyramid representation of shape features. The main advantages of our approach include the use of multiresolution descriptions to model object parts at different scales and the flexibility in handling partially intersected symbols in the matching process. Future work will focus on seeking effective methods to improve the efficiency, such as reducing the feature dimensionality using vector quantization-based techniques or avoiding explicit point-to-point matching by various probabilistic clustering or voting frameworks.

## ACKNOWLEDGMENT

Research supported by the National Science Foundation of China under Grant Nos. 61003113, 61021062 and the Natural Science Foundation of Jiangsu Province of China under Grant Nos. BK2009082 and the 973 Program of China under Grant No. 2010CB327903.

## REFERENCES

[1] K. Tombre, S. Tabbone, and P. Dosch, "Musings on symbol recognition," in *Graphics Recognition: Ten Year Review and Perspectives (GREC 2005)*, W. Liu and J. Llados, Eds., 2006, pp. 23–34.

[2] J. Llados, E. Valveny, G. Sanchez, and E. Mart, "Symbol recognition: Current advances and perspectives," in *Graphics Recognition Algorithms and Applications (GREC 2002)*, D. Blostein and Y.-B. Kwon, Eds., 2002, pp. 104–128.

[3] S. Yang, "Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor," *IEEE Trans. PAMI*, vol. 27, no. 2, pp. 278–281, Feb. 2005.

[4] W. Zhang, L. Wenyin, and K. Zhang, "Symbol recognition with kernel density matching," *IEEE Trans. PAMI*, vol. 28, no. 12, pp. 2020–2024, Dec. 2006.

[5] A. D. Bimbo and P. Pala, "Shape indexing by multi-scale representation," *Image and Vision Computing*, vol. 17, p. 245C261, 1999.

[6] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *ICCV 2005*, 2005, pp. 1458–1465.

[7] B. Li and H. Johan, "A comparison study on two multi-scale shape matching schemes," in *LNCS 5359 (ISVC 2008)*, G. B. et al., Ed., 2008, pp. 440–449.

[8] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. PAMI*, vol. 24, no. 24, pp. 509–522, Apr. 2002.

[9] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour and boundary detection improved by surround suppression of texture edges," *Image and Vision Computing*, vol. 22, no. 8, pp. 609–622, 2004.

[10] K. Grauman and T. Darrell, "Approximate correspondences in high dimensions," in *NIPS 19*, 2007, pp. 505–512.

[11] "International symbol recognition contest GREC'2005," http://symbcontestgrec05.loria.fr/.