A Comparison Study on Two Multi-scale Shape Matching Schemes

Bo Li and Henry Johan

School of Computer Engineering, Nanyang Technological University
Block N4, Nanyang Avenue, 639798, Singapore
{libo0002,henryjohan}@ntu.edu.sg

Abstract. We present and compare two multi-scale shape matching schemes: Chi-square distance based scheme and pyramid matching mode based scheme. We define a shape as a set of points. Multi-scale shape matching includes two steps: multi-scale feature extraction and point correspondence. We define a hybrid feature for every point by combining a global multi-scale shape context feature and a local variation feature. The two schemes have a difference in the computation of multi-scale shape context feature distance: the Chi-square distance based scheme directly sums up weighted Chi-square distances at different scales while the pyramid matching mode based scheme utilizes a multi-scale pyramid matching mode. Experimental results based on Frenkel and Kimia databases show that: (1) the pyramid matching mode based scheme can achieve robust and often better performance than the Chi-square distance based scheme; (2) the proposed two multi-scale schemes can achieve averagely better results than the single scale schemes.

1 Introduction

Shape matching is one of the key components for many computer vision and multimedia content retrieval problems. It includes two steps: feature extraction and point correspondence computation. Shapes have many representations such as point sets, curves and regions [17]. In this paper, we concentrate on point sets. Generally, the representations of a shape are highly dependent on the scales. For example, a tree may be depicted as a combination of trunks, branches and leaves, or only some points in a picture when we choose different scales. Multi-scale analysis can be used to extract interesting structures of a shape. Considering these, our work is dedicated to finding effective shape matching schemes by adopting a multi-scale strategy.

We propose two multi-scale shape matching schemes which are Chi-square distance based scheme and pyramid matching mode based scheme. We define a hybrid feature of a point as the combination of a global feature shape context [2] and a local variation feature [10, 14]. Another contribution of our work is that we present two novel schemes to compute the multi-scale shape context feature distance. The Chi-square distance based scheme defines the multi-scale shape context feature distance by directly summing up weighted Chi-square feature distances at different scales. The pyramid matching mode based scheme utilizes a multi-scale pyramid matching mode to compute the distance. To compute the correspondences between two point sets, we

G. Bebis et al. (Eds.): ISVC 2008, Part II, LNCS 5359, pp. 440–449, 2008.

adopt fast marching method (FMM) [4] and Jonker's linear assignment problem (LAP) method [8]. Experimental results based on Frenkel [4] and Kimia [2] databases show that the pyramid matching mode based scheme has a more robust and better performance than the Chi-square distance based scheme and the proposed two multi-scale schemes can achieve averagely better results than the single scale schemes.

The paper is organized as follows. We briefly discuss related work in Section 2. Section 3 introduces the shape context and local variation features. In Section 4, we describe our proposed two multi-scale shape matching schemes. We give in details the experiments and comparison of the two multi-scale shape matching schemes in Section 5. We conclude the work in Section 6.

2 Related Work

A survey on general shape matching is presented in [11]. An overview of shape matching techniques based on computer geometry is given in [17]. We focus on work closely related to the two important facets of our multi-scale shape matching schemes: shape features selection and multi-scale feature extraction.

Selecting an appropriate feature is very important in shape matching. Shape features can be classified into local and global features. Some commonly used features are shape context [2], local variation [10, 14], curvatures, tangent vectors, salient geometric features [6] and key points or corners [16]. A hybrid feature is a combination of two or more such shape features. Its main objective is to make the resulting hybrid feature has more properties and the components of the hybrid feature complement with each other.

Multi-scale feature extraction means using different scales to characterize the same feature. It should be differentiated from multi-resolution feature extraction which usually adopts some multi-resolution analysis or representation tools such as wavelet transform [1, 3]. Among the multi-scale shape matching methods, pyramid matching has attracted more and more interests in recent years. Grauman and Darrell [7] presented a kernel-based pyramid matching algorithm and a scheme to compute similarity between histograms. Lazebnik et. al [9] presented a spatial pyramid matching framework based on histogram analysis on sub-regions of an image. Maji et. al [12] verified that classification using intersection kernel support vector machine can be more efficient than the standard approaches after using the histogram intersection kernel.

3 Global and Local Features

Shape Context. The descriptor of shape context (2D) is first proposed by Belongie et. al [2] and it is a rich global feature descriptor that has been successfully used in various application fields [2,4,5,13]. It divides a point's surrounding area into several bins with uniformly increasing distances and angles in log-polar space. The shape context of a point P_i is defined by computing the percentages of other points that are located in each of its surrounding bins,

$$h_i(k,l) = \#\{x \neq P_i : (x - P_i) \in \text{bin}(k,l)\} / X \ . \tag{1}$$

X is the total number of points. k and l are the distance and orientation bin index respectively. In other words, a point's shape context feature is actually a log-polar histogram which defines the relative distribution of other points. Different points in one shape have different shape context features and similar points in two similar shapes have similar shape context features. Paper [2] adopts the Chi-square distance to measure the difference between the shape context features of two points P_i and P_j.

$$C_{ij} = \frac{1}{2} \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{[h_i(k,l) - h_j(k,l)]^2}{h_i(k,l) + h_j(k,l)} \ . \tag{2}$$

h_i and h_j are the normalized shape context histograms of P_i and P_j. C_{ij} is in the range of [0,1]. To achieve rotation invariance, the author proposed a relative frame-based shape context feature by adopting the local tangent vector at each point as the reference axis for angle computation. But the shape context feature using relative frame will also partially lose its ability of discrimination. For example, it is difficult to distinguish "6" between "9". Therefore, we propose a hybrid feature combining the global shape context feature and the rotation invariant feature local variation.

Local Variation. Local variation [10, 14] is a local feature using covariance analysis on a point's local neighborhood. For a 2D shape, the covariance matrix has two eigenvalues λ_N and λ_T. The eigenvalue λ_N is to measure the extent of the neighboring points' deviation to the normal direction and λ_T is to measure the variation of the neighboring points' distribution in the tangent direction [14]. Ligang et. al. [10] defined a local variation feature of point P_i as follows,

$$\sigma(P_i, \delta) = \varepsilon \frac{\lambda_N}{\lambda_N + \lambda_T} \ . \tag{3}$$

δ is the radius of the neighborhood. When the neighborhood is a convex, ε=1; when it is a concave, ε=-1. $\sigma(P_i,\delta) \in$ [-1,1]. Local variation $\sigma(P_i,\delta)$ only depends on the relative distribution of its neighborhood, therefore it is a rotation invariant feature. The difference between the local variations of two points can be defined as follows.

$$L_{ij} = \left| \sigma(P_i, \delta_1) - \sigma(P_j, \delta_2) \right| / 2 \ . \tag{4}$$

δ_1 and δ_2 are the radii of the neighborhoods of point P_i and point P_j respectively. L_{ij} is also in the range of [0,1].

4 Multi-scale Shape Matching Schemes

In this part, we present two multi-scale matching schemes. Assume that the two shapes I_1 and I_2 have m and n points respectively. Our matching algorithms adopt the following steps: first extract the features of the points and then find a metric to represent the distance of two points' features and after that we get an m x n distance matrix. Then the next thing to do is to find an optimal correspondence (or assignment) method to make the total distance of all the points in the two shapes be minimal. In details, the steps of the two schemes are as follows.

(1) Construct shape context feature scale-spaces H_1 and H_2 for the two shapes (See details in Section 4.1).

(2) Compute multi-scale shape context distance matrix *S* (See details in Section 4.2, Subsection Scheme 1 for the Chi-square distance based scheme and Subsection Scheme 2 for the pyramid matching mode based scheme).

(3) Compute local variation distance matrix *L* (See details in Section 4.3 for both schemes).

(4) Combine multi-scale shape context distance matrix *S* with local variation distance matrix *L* together by the feature weights computed based on their discrimination ability to formulate a hybrid multi-scale distance matrix *F* (See details in Section 4.4 for both schemes).

(5) Matching or correspondence using fast marching method (FMM) [4] or Jonker's linear assignment problem (LAP) [8].

We can see that the two schemes only have a difference in the second step.

4.1 Construction of Shape Context Feature Scale-Spaces

Belongie et. al [2] and Frenkel et. al [4] both used a fixed type of shape context (5x12, 5 distance bins and 12 orientation bins). This means that the numbers of distance and orientation bins remain unchanged during the process of matching. However, in these work the authors did not state the method of deciding their appropriate values. Through experiments, we found that the matching results are more or less different when using different numbers of the distance and orientation bins. It is because the result of analysis of a shape is highly dependent on the scale selected. To address this issue, we organize a shape context scale-space by dividing a point's surrounding area into different numbers of distance bins N_r and orientation bins N_o (not fix to 5x12) during matching based on exponential division: N_r=N_o=2^s, *s* is the scale. Assume that the scale *s* varies from the minimum scale $s_{\min}$ to the maximum scale $s_{\max}$. For each scale we can compute shape context feature for every point using Equation (1) and then construct shape context feature scale-spaces H_1 and H_2 for the two shapes I_1 (*m* points) and I_2 (*n* points) as follows,

$$H_1 = \{h_i^s \mid s_{\min} \le s \le s_{\max}, 1 \le i \le m\}, \qquad H_2 = \{h_j^s \mid s_{\min} \le s \le s_{\max}, 1 \le j \le n\}. \tag{5}$$

h_i^s and h_j^s are the normalized shape context histograms of a point P_i in I_1 and another point Q_j in I_2 at the scale *s*.

4.2 Computation of Multi-scale Shape Context Distance Matrix

Scheme 1 — Chi-square Distance Based Method. The steps are as follows.

(1) Compute shape context feature distance matrix C^s (*m* x *n*) on each scale *s* ($s_{\min} \le s \le s_{\max}$) based on the two feature scale-spaces H_1 and H_2. The shape context distance between a point P_i in I_1 and another point Q_j in I_2 is as follows,

$$C_{ij}^{s}=\frac{1}{2}\sum_{k=1}^{N_r}\sum_{l=1}^{N_O}\frac{[h_i^s(k,l)-h_j^s(k,l)]^2}{h_i^s(k,l)+h_j^s(k,l)}\ . \tag{6}$$

$1 \le i \le m$, $1 \le j \le n$, $N_r = N_o = 2^s$ and k, l are the distance and orientation bin index.

(2) Weight the shape context distance matrices at different scales together to get the multi-scale shape context distance matrix S (m x n).

$$S = \sum_{s=s_{\min}}^{s_{\max}} w_s C^s, \qquad w_s = \frac{1}{2^{s_{\max}-s}}\ . \tag{7}$$

Scheme 2 — Pyramid Matching Mode Based Method. This scheme is triggered by the pyramid matching method proposed in [7]. The pyramid matching method first maps two feature sets to two multi-scale histogram pyramids based on a new kernel function. Then, it computes the intersections of two feature sets' histograms on each level of the pyramids. Finally, it weights the difference between successive levels' intersection together to get the similarity of the two feature sets. The kernel entitles us to transfer matching result from a lower level to an adjacent higher level.

In fact, this pyramid matching method can be viewed as a multi-scale matching scheme too. We can utilize this framework to formulate another scheme to compute the shape context feature distance matrix and it is very different from Chi-square distance based scheme. However, we need to make some adjustments since there are two important differences between the pyramid matching method and our approach.

The first key difference is that we do not need to construct histogram pyramids using the kernel function described in [7] since we can use the multi-scale shape context features directly after constructing feature scale-spaces (See Section 4.1). In fact, our multi-scale shape context features are organized based on the numbers of distance and orientation bins rather than the different sizes of histogram like [7].

Second, we want to get the distance rather than the similarity of two shapes. Therefore the definition of intersection of two shapes' feature histograms in our algorithm is also different (See Equation (8)).The complete steps are as follows.

(1) Compute the histogram intersection matrix H^s (m x n) for the two feature scale spaces. The histogram intersection between a point P_i in I_1 and another point P_j in I_2 on the scale s is as follows,

$$H_{ij}^{s}=1-\sum_{k=1}^{N_r}\sum_{l=1}^{N_o} min(h_i^s(k,l),h_j^s(k,l))\ . \tag{8}$$

$1 \le i \le m$, $1 \le j \le n$, $N_r = N_o = 2^s$, $s_{\min} \le s \le s_{\max}$. k, l are the distance and orientation bin index.

(2) Compute the difference between successive levels' intersections.

$$N^s = H^s - H^{s-1}, \qquad s_{\min}+1 \le s \le s_{\max}\ . \tag{9}$$

(3) Combine the difference of each scale with a weight value related to the scale value to formulate the multi-scale shape context feature distance matrix S.

$$S = \sum_{s=s_{\min}+1}^{s_{\max}} w_s N^s, \qquad w_s = \frac{1}{2^{s_{\max}-s}}. \tag{10}$$

4.3 Computation of Local Variation Feature Distance Matrix

The steps are as follows.

(1) Compute the local variation feature arrays V_1 and V_2 for the two shapes. If choosing different values for neighborhood threshold δ, we can get different local variation feature σ^δ. Assume $\delta_{\min}$ and $\delta_{\max}$ are the minimum and maximum values for δ. First, compute each local variation feature value for each point using Equation (3). At certain scale the local variation feature has a local extreme [15] which represents a relatively apparent as well as our interesting local variation feature σ_i,

$$\sigma_i = \max_{\delta}\{\sigma_i^\delta \,|\, \delta_{\min} \le \delta \le \delta_{\max}\}. \tag{11}$$

Then, compute the local variation feature arrays V_1 and V_2 for the two shapes I_1 (m points) and I_2 (n points) as follows,

$$V_1 = \{\sigma_i \mid 1 \le i \le m\}, \qquad V_2 = \{\sigma_j \mid 1 \le j \le n\}. \tag{12}$$

(2) Compute the local variation feature distance matrix L (m x n) using Equation (4) for V_1 and V_2.

4.4 Formulation of Hybrid Multi-scale Feature Distance Matrix

First, compute the weights w_S and w_L for the two features shape context and local variation respectively,

$$w_S = \frac{\sigma_S}{\sigma_S + \sigma_L}, \qquad w_L = \frac{\sigma_L}{\sigma_S + \sigma_L}. \tag{13}$$

σ_S and σ_L are the standard deviations of the matrices S and L respectively. Usually the standard deviation of a feature is a measure of its dispersion property showing its differentiation ability. Therefore our above proposed automatic weight assignment method indicates giving an appropriate weight to a feature according to its differentiation.

Next, combine the shape context and local variation feature distance matrix using the weight values to formulate a hybrid multi-scale feature distance matrix F,

$$F = w_S \cdot S + w_L \cdot L. \tag{14}$$

5 Experiments and Discussion

In this part, we compare the two schemes' matching performance and characteristics such as error rate and robustness based on the two databases in [2] and [4].

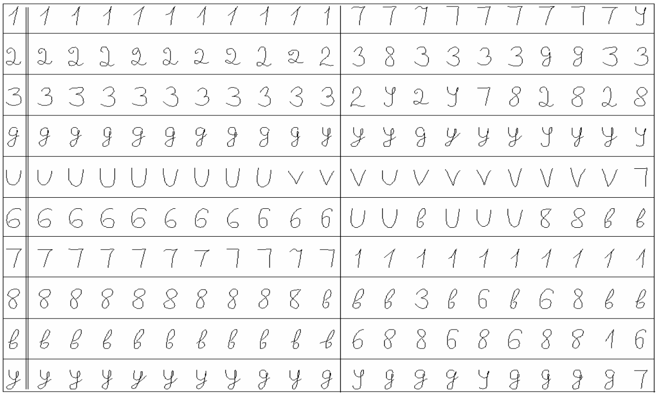

Fig. 1. 20-best matches example for Frenkel database using SC2

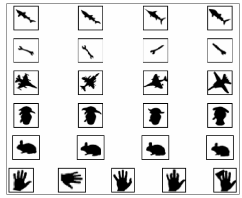

Fig. 2. Kimia database

5.1 Frenkel Curves Database Matching

Frenkel et. al [4] created a curves database comprising 110 curves, categorized into 11 types, 10 each. The first 10 prototype curves are shown in the first column of Figure 1. We use our presented two schemes to do a complete matching: each time select one prototype from the first 10 prototypes and match it with the whole curves database. One factor to be considered is the scales s ($s_{min} \leq s \leq s_{max}$). Larger s gets more accurate shape context feature, but needs more time. We choose three multi-scale combinations: <s_{min}=2, s_{max}=3>, <s_{min}=3, s_{max}=4> and <s_{min}=4, s_{max}=5>. We denote the Chi-square distance based scheme as SC1, and the pyramid matching mode based scheme as SC2. In order to compare the multi-scale and the single scale matching schemes, we use the single scale shape context (not 5x12 but 2^sx2^s) distance similar to [4] to replace the multi-scale shape context distance in the second step of the two schemes (Section 4.2). We denote this single scale scheme as SC_F. Figure 1 gives an example of the 20-best matching results of SC2 on the scales <s_{min}=3, s_{max}=4>. Table 1 shows the comparison of the three methods. *R* is the recognition rate.

Table 1. 10-best matches results for the three methods

Scales \ Methods	SC_F (s)	SC_F (R)	SC1 (R)	SC2 (R)
<s_{min}=4, s_{max}=5>	5	88%	90%	92%
	4	95%		
<s_{min}=3, s_{max}=4>	3	94%	95%	94%
<s_{min}=2, s_{max}=3>	2	93%	94%	94%

First, as seen from Table 1, SC2 has more moderate fluctuations than SC1 over the three types of scale combinations. SC1 may have a disadvantage of instability. For instance, the recognition rate on the scales <s_{min}=4, s_{max}=5> is only 90%.

Second, as shown in Table 1, one advantage of the multi-scale scheme SC2 over the single scale scheme SC_F is that its performance is not highly dependent on the scale selection. This is because of its principle for computing the multi-scale features. It is based on the difference between successive levels, not on the absolute feature values on each level. SC2 can achieve good results more easily than SC_F. As shown in Table 1, with only very coarse scales such as <s_{min}=2, s_{max}=3>, SC2 already gets a result among the top best ones. But, for SC_F we find that finer scale (for example, s=5) does not mean more accurate result (only 88%). Therefore, it is more difficult to choose an appropriate scale to achieve better results for SC_F than for SC2. The average recognition rates for SC_F, SC1 and SC2 are 92.5%, 93% and 93.3%. The two multi-scale schemes achieve better results than the single scheme averagely.

In order to know whether the local variation feature really works as an important part of the hybrid feature, we also do comparison experiments for the two schemes with and without local variation feature. We denote the two schemes without local variation feature as SC1_S and SC2_S. The results are shown in Table 2.

Table 2. 10-best matches results for the two schemes with and without local variation feature

Methods / Scales	SC1_S	SC1	SC2_S	SC2
<s_{min}=4, s_{max}=5>	92%	90%	92%	92%
<s_{min}=3, s_{max}=4>	94%	95%	94%	94%
<s_{min}=2, s_{max}=3>	95%	94%	92%	94%

We find that SC2 improves more than SC1 after adding the local variation feature. For example, SC2 improves the recognition rate by 2% after adding the local variation feature at the coarse scales <s_{min}=2, s_{max}=3>. On the contrary, SC1 deteriorates at a coarser (<s_{min}=2, s_{max}=3>) and finer scale (<s_{min}=4, s_{max}=5>) and only has an improvement of 1% on the middle scale (<s_{min}=3, s_{max}=4>).

5.2 Kimia Database Matching

Kimia Database [2] has 25 images and 6 categories (see Figure 2). We use the first column as the prototypes and do the experiment similar to Section 5.1. We can see rotation operation has been applied to some images of the database. Therefore, we need to adopt the relative shape context feature as described in Section 3.1. The selections of multi-scale combinations are the same as those in Section 5.1.We first extract the contour of each image and use cubic spline interpolation to represent the contour curve and then uniformly sample 100 points. We use Jonker's LAP method as the corresponding algorithm. We also use the single scale shape context distance similar to [2] to replace the multi-scale shape context distance and denote this single scale scheme as SC_B. Table 3 shows the 3-best (exclude the prototype itself) matching results for the three schemes, and the three digits in the bracket means the number of images that are correctly classified into its belonging types for the 3-best matches respectively.

Table 3. 3-best matches results for Kimia database matching

Scales \ Methods	SC_B (s)	SC_B (n_1, n_2, n_3)	SC1 (n_1, n_2, n_3)	SC2 (n_1, n_2, n_3)
<s_{min}=4, s_{max}=5>	5	(20,18,13)	(23,18,12)	(25,17,18)
	4	(24,19,13)		
<s_{min}=3, s_{max}=4>	3	(24,19,12)	(24,18,14)	(24,17,16)
<s_{min}=2, s_{max}=3>	2	(19,18,11)	(24,18,14)	(22,20,13)

Table 4. 3-best matches results for the two schemes with and without local variation feature

Scales \ Methods	SC1_S	SC1	SC2_S	SC2
<s_{min}=4, s_{max}=5>	(23,20,12)	(23,18,12)	(24,19,12)	(25,17,18)
<s_{min}=3, s_{max}=4>	(22,19,10)	(24,18,14)	(23,19,10)	(24,17,16)
<s_{min}=2, s_{max}=3>	(22,17,11)	(24,18,14)	(22,17,12)	(22,20,13)

Results indicate that SC2 outperforms SC1 and SC_B on the average and their average recognition rates (total number of correctly classified images/total number of matching) are 76.4%, 73.2% and 70.0% respectively. We can also conclude that the two multi-scale schemes achieve better results than the single scheme averagely.

Then, we compare the performances of the two schemes with and without the local variation feature, as shown in Table 4. The average recognition rates of SC1_S and SC2_S are 69.3% and 70.2%. We also find SC2 has a more apparent improvement (6.2%) than SC1 (3.9%) does after adding a local variation feature.

6 Conclusion

The contributions of this paper are as follows. First, we define a hybrid feature which combines a global multi-scale shape context feature and a local variation feature for shape matching. Second, we provide two multi-scale shape matching schemes by adopting different multi-scale feature distance computing methods. The Chi-square distance based scheme computes the multi-scale shape context feature distance by directly summing up a point's weighted Chi-square feature distances at different scales while the pyramid matching mode based scheme utilizes a multi-scale pyramid matching mode to implement this. Third, we do a comparison study on them based on two data sets and find that the pyramid matching mode based scheme can achieve a more robust and often better result than the Chi-square distance based scheme. In addition, the proposed two multi-scale schemes can achieve averagely better results than the single scale schemes.

References

1. Antoine, J.P., Barache, D., Cesar, J.R.M., Costa, D.F.: Shape Characterization with the Wavelet Transform. Signal Processing 62(3), 265–290 (1997)
2. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. IEEE Transaction on Pattern Analysis and Machine Intelligence 24(24), 509–522 (2002)
3. Finkelstein, A., Salesin, D.H.: Multiresolution Curves. Computer Graphics 28(4), 261–268 (1994)
4. Frenkel, M., Basri, R.: Curve Matching Using the Fast Marching Method. In: Proceedings of Energy Minimization Methods in Computer Vision and Pattern Recognition, 4th International Workshop, Lisbon, Portugal, pp. 35–51 (2003)
5. Frome, A., Huber, D., Kolluri, R., Bulow, T., Malik, J.: Recognizing Objects in Range Data Using Regional Point Descriptors. In: Proceedings of the European Conference on Computer Vision, pp. 224–237 (2004)
6. Gal, R., Cohen-Or, D.: Salient Geometric Features for Partial Shape Matching and Similarity. ACM Transactions on Graphics 25(1), 130–150 (2006)
7. Grauman, K., Darrell, T.: The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features. In: IEEE International Conference on Computer Vision, vol. 2, pp. 1458–1465 (2005)
8. Jonker, R., Volgenant, A.: A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems. Computing 38(4), 325–340 (1987)
9. Lazebnik, S., Schmid, C., Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178 (2006)
10. Ligang, L., Guopu, W., Bo, Z., Baining, G., Heung-Yeung, S.: Perceptually Based Approach for Planar Shape Morphing. In: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference, pp. 111–120 (2004)
11. Loncaric, S.: A Survey of Shape Analysis Techniques. Pattern Recognition 31(8), 983–1001 (1998)
12. Maji, S., Berg, A.C., Malik, J.: Classification Using Intersection Kernel Support Vector Machines Is Efficient. IEEE Computer Vision and Pattern Recognition (to appear, 2008)
13. Mori, G., Malik, J.: Estimating Human Body Configurations Using Shape Context Matching. In: Proceedings of the 7th European Conference on Computer Vision, pp. 666–680 (2002)
14. Pauly, M., Markus, G., Leif, P.K.: Efficient Simplification of Point-Sampled Geometry. In: Proceedings of the Conference on Visualization, pp. 163–170 (2002)
15. Pauly, M., Keiser, R., Gross, M.: Mutli-Scale Feature Extraction on Point-Sampled Surfaces. In: Eurographics 2003, pp. 281–289 (2003)
16. Smith, S.M., Brady, J.M.: SUSAN—A New Approach to Low Level Image Processing. International Journal of Computer Vision 23(1), 45–78 (1997)
17. Veltkamp, R.C., Hagedoorn, M.: State-of-the-art in Shape Matching. Technical Report UU-CS-1999-27, Utrecht University, the Netherlands (1999)