Shape-based Object Matching Using Point Context

Cong Yang, Christian Feinen, Oliver Tiebe, Kimiaki Shirahama and Marcin Grzegorzek
Research Group for Pattern Recognition
University of Siegen, Germany
Hoelderlinstr. 3, D-57076 Siegen
{cong.yang,marcin.grzegorzek}@uni-siegen.de

ABSTRACT

This paper proposes a novel object matching algorithm based on shape contours. In order to ensure low computational complexity in shape representation, our descriptor is composed by a small number of interest points which are generated by considering both curvatures and the overall shape trend. To effectively describe each point of interest, we introduce a simple and highly discriminative point descriptor, namely Point Context, which represents its geometrical and topological location. For shape matching, we observed that the correspondences are not only dependent on the similarities between these single points in different objects, but they are also related to the geometric relations between multiple points of interest in the same object. Therefore, a high-order graph matching formulation is introduced to merge the single point similarities and the similarities between point triangles. The main contributions of this paper include (i) the introduction of a novel shape descriptor with robust shape points and their descriptors and (ii) the implementation of a high-order graph matching algorithm that solves the shape matching problem. Our method is validated through a series of object retrieval experiments on four datasets demonstrating its robustness and accuracy.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis; I.5.4 [**Pattern Recognition**]: Computer Vision; I.5.3 [**Pattern Recognition**]: Similarity Measures; F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Pattern Matching

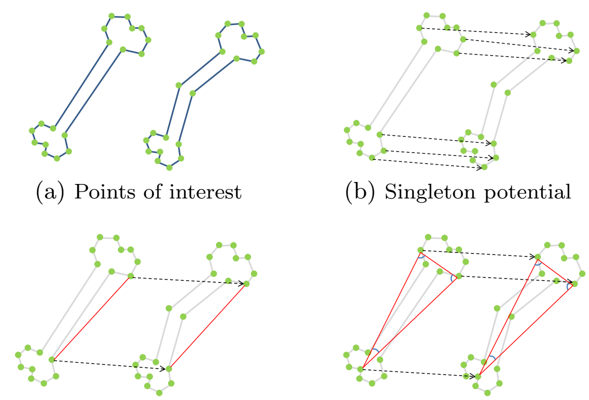
General Terms

Experimentation, Performance

Keywords

Shape matching; Interest point detection; Point context; High-order graph matching

ICMR '15, June 23 - 26, 2015, Shanghai, China
Copyright 2015 ACM 978-1-4503-3274-3/15/06 ...$15.00.
http://dx.doi.org/10.1145/2671188.2749329



(a) Points of interest (b) Singleton potential

(c) Pairwise potential (d) Third-order potential

Figure 1: Different potentials for object matching.

1. INTRODUCTION

Object recognition is a process for identifying a specific object in a digital image or video, useful for a number of applications. Shape analysis plays a key role in understanding or identifying objects in images [11, 13]. Contour-based shape matching generally consists of two main processes, shape representation and matching.

For the shape representation process, there are three main challenges. The first challenge is how to extract efficient descriptors that are invariant to shape rotation, translation and scaling. The second one is how to extract appropriate shape descriptors in images that are often noisy and/or distorted. The third challenge is how to generate shape descriptors with low computation complexity. In order to solve these problems we represent the shape only with a small number of interest points. We detect these points by considering both curvatures and the overall shape trend. Secondly, there is only a limited number of these points in each object, which can reduce the computational complexity for searching correspondences dramatically.

For the shape matching process, even using only the proposed points of interest, several points are inevitably extracted regardless of their irrelevance to the matching process (the right bone in Figure 1(a)). Moreover, some extracted points in the same object may have very similar geometrical locations because of their small distances or symmetry. However, most existing methods [10, 1, 6] only consider the relationship between single points in different objects. This could lead to ambiguous matching because many different points may have similar descriptors [5, 13]. It has been shown that high-order graph matching is a powerful framework to establish feature correspondences, combining both appearance similarity and geometric compatibility [12]. As shown in Figure 1(b), singleton point matching is a well-know assignment problem. For the pairwise matching (Figure 1(c)), it finds consistent correspondences between two sets of points, by taking in consideration both how well the features' descriptor match and how well their pairwise geometric constraints are satisfied. For the high-order matching (mostly third-order, see Figure 1(d)), it considers the cost of matching three correspondences. With this observation, we present a high-order graph matching strategy for improving correspondences between our points of interest.

2. SHAPE DESCRIPTOR

In this section, we first describe the method that generates robust points of interest along the shape boundary. After that, the point descriptor, namely point context, is introduced and analysed. Finally, we construct the shape descriptor with the interest points and their associated point context features.

2.1 Points of Interest

In this section an approach [5] is employed on an arbitrary given shape Ω to select boundary points $\boldsymbol{p}_i$ which are significant for the perceptual appearance. The general idea is to detect contour regions like the legs or the tail of an elephant characterised by a higher curvature toward the overall shape trend. Therefore, either a single or multiple *reference points* ($\boldsymbol{x}_i$) inside the shape are determined to compute the distance between each single contour point and its closest reference. By arranging these values sequentially, a signal $\boldsymbol{s}$ is generated that can be used to detect peaks. The desired subset of contour points is then obtained by assigning the corresponding contour point to each peak.

2.2 Point Context

Inspired by [9, 8], for points of interest $\boldsymbol{p}_i, (i = 1, \dots, m)$, we propose a point descriptor, Point Context, which can intuitively capture its geometrical and topological locations. We consider the set of vectors originating from p_i to all other sample points on a shape contour. These vectors express the configuration of the entire shape relative to the reference point. More specifically, let P denote a sequence of interest points $P = \{\boldsymbol{p}_1, \cdots \boldsymbol{p}_m\}$ and Q denote a finite number of contour sample points $Q = \{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\}$, $P \notin Q$. All points in P and Q are represented by their coordinate locations. Points in Q are ordered clockwise along the shape contour. For each single point of interest $\boldsymbol{p}_i$, we compute two vectors, one presenting all distances and the second representing all pairwise orientations of vectors from $\boldsymbol{p}_i$ to each sample point $\boldsymbol{q}_k \in Q$, $(k = 1, \cdots, n)$. A distance $D^{\boldsymbol{p}_i}(k)$ from $\boldsymbol{p}_i$ to $\boldsymbol{q}_k$ is defined as Euclidean distance in the log space

$$D^{\boldsymbol{p}_i}(k) = \log(1 + \|\overrightarrow{\boldsymbol{p}_i} - \overrightarrow{\boldsymbol{q}_k}\|^2) \quad . \tag{1}$$

In order to avoid the situation where the input of log is zero, we add one to Euclidean distance. An orientation $\Theta^{\boldsymbol{p}_i}(k)$ from $\boldsymbol{p}_i$ to $\boldsymbol{q}_k$ is defined as the orientation of vector $\overrightarrow{\boldsymbol{p}_i} - \overrightarrow{\boldsymbol{q}_k}$:

$$\Theta^{\boldsymbol{p}_i}(k) = \text{atan2}(\overrightarrow{\boldsymbol{p}_i} - \overrightarrow{\boldsymbol{q}_k}) \quad . \tag{2}$$

where atan2 stands for the four quadrant inverse tangent which can ensure $\Theta^{\boldsymbol{p}_i}(k) \in [-\pi, \pi]$. Together with the distances, a single point of interest $\boldsymbol{p}_i$ is encoded in two n-dimensional vectors $D^{\boldsymbol{p}_i}$ and $\Theta^{\boldsymbol{p}_i}$.

Our proposed point descriptor differs from the methods [5] and [3] in the following aspects. Firstly, we only consider the feature vectors from points of interest instead of roughly uniform spacing or randomly picking for selecting sample points. This strategy can reduce the mismatches and computational complexity conspicuously. Secondly, the proposed point descriptor is naturally translation and rotation invariant [1]. However, approaches in [5, 3] are not intrinsically rotation invariant. In order to solve this problem, they use the tangent angle on each point to turn the shape. This is not robust enough for some points which have no reliable tangents. Moreover, some local appearance features could lose their discriminative power if the shape is rotated.

Eventually, based on our method, for a given arbitrary shape Ω, its contour $\partial\Omega$ can be represented with the locations as well as the distance and orientation vectors of all points of interest from the contour:

$$\partial\Omega = \{\boldsymbol{p}_i, D^{\boldsymbol{p}_i}, \Theta^{\boldsymbol{p}_i}\} \quad . \tag{3}$$

3. SHAPE MATCHING

In this section, we first formulate our shape matching based on the properties of interest points and point context features. The formulation is composed by potential functions with different orders. After that, we introduce the definitions of potential functions.

3.1 Formulation of Shape Matching

Let P_1 and P_2 denote the set of interest points from two shapes S_1 and S_2 respectively. $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$ denote a single point of interest in P_1 and P_2 respectively. $P \triangleq P_1 \times P_2$ denotes the set of possible correspondences. We define the Boolean indicator variable

$$x_a = \begin{cases} 1 & \text{if } a = (\boldsymbol{p}_i, \boldsymbol{p}'_j) \in P \text{ is a correspondence} \\ 0 & \text{otherwise} \end{cases} \quad . \tag{4}$$

In our definition, a basic constraint is that each point $\boldsymbol{p}_i$ in P_1 is mapped to at most one point $\boldsymbol{p}'_j$ in P_2, while for each point $\boldsymbol{p}'_j$ in P_2 there is at most one point $\boldsymbol{p}_i$ in P_1 mapping to it. Therefore, we have the set of feasible solutions:

$$\begin{aligned} \zeta = \{\mathbf{x} \in \{0,1\}^{P_1 \times P_2} \mid \sum_{\boldsymbol{p}_i \in P_1} x_{\boldsymbol{p}_i, \boldsymbol{p}'_j} \leqslant 1, \sum_{\boldsymbol{p}'_j \in P_2} x_{\boldsymbol{p}_i, \boldsymbol{p}'_j} \leqslant 1, \\ \forall \boldsymbol{p}_i \in P_1 \text{ and } \forall \boldsymbol{p}'_j \in P_2\} \end{aligned} \quad . \tag{5}$$

Inspired by [4, 14], our high-order (degree 3) matching formulation is required to be as follows:

$$\begin{aligned} \min_{\mathbf{x} \in \zeta} \{E(\mathbf{x}|\theta) = \sum_{a \in P} \theta_a x_a + \sum_{(a,b) \in P \times P} \theta_{ab} x_a x_b + \\ \sum_{(a,b,c) \in P \times P \times P} \theta_{abc} x_a x_b x_c\} \end{aligned} \quad . \tag{6}$$

where θ_a is the matching cost for each correspondence $a \in P$ (Figure 1(b)), θ_{ab} for a pair of correspondences $(a, b) \in P \times P$ (Figure 1(c)), and θ_{abc} for a triplet of correspondences $(a, b, c) \in P \times P \times P$ (Figure 1(d)). Since the matching constraint in Eq. 5 makes the optimisation problem in Eq. 6 difficult to solve, we employ the dual-decomposition method [14] to re-formulate the original problem as the union of several sub-problems that are easier to solve.

[1]Cooperating with Eq. 7, our descriptor is also scale invariant.

3.2 Potential Functions

In this paper, we only consider the first and third order terms since a triplet (Figure 1(d)) in one shape would have fewer similar triplets in the other. Moreover, this strategy can be used to avoid the over-fitting problem as the singleton potentials still offer the major contribution for the shape similarity with our proposed point context descriptors.

3.2.1 The Singleton Potential

In order to define the singleton potential θ_a, for each correspondence $(\boldsymbol{p}_i, \boldsymbol{p}'_j)$ we fully use the properties of its point context features. Given two points of interest $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$ from two shapes S_1 and S_2, respectively, we first compute the affinity vectors between the corresponding elements in their distance and orientation vectors:

$$A_D(\boldsymbol{p}_i, \boldsymbol{p}'_j) = \exp(-\frac{(D^{\boldsymbol{p}_i}(k) - D^{\boldsymbol{p}'_j}(k))^2}{(\max(D^{\boldsymbol{p}_i})\sigma)^2}) \quad . \tag{7}$$

$$A_\Theta(\boldsymbol{p}_i, \boldsymbol{p}'_j) = \exp(-\frac{(\Theta^{\boldsymbol{p}_i}(k) - \Theta^{\boldsymbol{p}'_j}(k))^2}{\delta^2}) \quad . \tag{8}$$

where σ and δ represent the tolerance of distance and orientation differences, respectively. We set $\sigma = 0.2$ and $\delta = \pi/4$ in all experiments. Like in Eq. 1, $k = 1, \cdots, n$. To get a scale invariant value of $A_D(\boldsymbol{p}_i, \boldsymbol{p}'_j)$, we divide each distance difference by the maximum of the first distance vector.

Since both A_D and A_Θ are normalised, we can simply add them to obtain the affinity vector:

$$A(\boldsymbol{p}_i, \boldsymbol{p}'_j) = A_D(\boldsymbol{p}_i, \boldsymbol{p}'_j) + A_\Theta(\boldsymbol{p}_i, \boldsymbol{p}'_j) \quad . \tag{9}$$

Observing that A is a $n-$dimensional vector representing the similarities of corresponding pairs from $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$. The similarity between $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$ can be calculated by the mean value of $A(\boldsymbol{p}_i, \boldsymbol{p}'_j)$. Consequently, the singleton potential for a correspondence $(\boldsymbol{p}_i, \boldsymbol{p}'_j)$ is defined as

$$\theta_a = \theta_{\boldsymbol{p}_i, \boldsymbol{p}'_j} = \frac{1}{n}\sum_{k=1}^{n} A(k) \quad . \tag{10}$$

3.2.2 The High-Order Potential

In this paper, we treat the third-order (triangle in Figure 1(d)) as the high-order potentials. The angles of a triangle are scale and rotation invariant. Thus we can describe each triangle by its three angles. In practice, we describe each triangle by the sine of its three angles to speed-up the computation. Suppose that P_1 and P_2 are the set of interest points for two shapes S_1 and S_2, respectively. For any two triplets, $(\boldsymbol{p}^1_i, \boldsymbol{p}^1_j, \boldsymbol{p}^1_k) \in S_1$ and $(\boldsymbol{p}^2_i, \boldsymbol{p}^2_j, \boldsymbol{p}^2_k) \in S_2$, the triple potential for each possible triple matching $(\boldsymbol{p}^1_i, \boldsymbol{p}^1_j, \boldsymbol{p}^1_k) \rightarrow (\boldsymbol{p}^2_i, \boldsymbol{p}^2_j, \boldsymbol{p}^2_k)$ is defined with a truncated Gaussian kernel:

$$\theta_{abc} = \theta_{\boldsymbol{p}^1_i, \boldsymbol{p}^1_j, \boldsymbol{p}^1_k, \boldsymbol{p}^2_i, \boldsymbol{p}^2_j, \boldsymbol{p}^2_k} = \begin{cases} \exp(-\gamma\|f_{i_1,j_1,k_1} - f_{i_2,j_2,k_2}\|^2) & \text{if } \|f_{i_1,j_1,k_1} - f_{i_2,j_2,k_2}\| \leqslant \vartheta \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where f_{i_1,j_1,k_1} and f_{i_2,j_2,k_2} are the feature vectors (sine) describing the triplets $(\boldsymbol{p}^1_i, \boldsymbol{p}^1_j, \boldsymbol{p}^1_k)$ and $(\boldsymbol{p}^2_i, \boldsymbol{p}^2_j, \boldsymbol{p}^2_k)$, respectively. Points in both triplets are ordered in a clockwise fashion with the singleton potential pair as its starting point $\boldsymbol{p}^1_i$ and $\boldsymbol{p}^2_i$. We use the truncated Gaussian kernel to scatter and reduce matching times since the number of possible triple matching is huge and it is not necessary to compute them completely. In this paper, we take $\gamma = 2$ in our experiment. With Eq. 11, for each tuple i of P_1, we find the features of P_2 in a neighbourhood of size ϑ ($\vartheta = 350$ in our experiment). Based on [4] we only sample 20 triangles per points in P_1 since this number of triangles is more than enough to obtain a robust matching. Then, we sample all the possible triangles of P_2, and compute their descriptors. We employ a kd-tree to store them efficiently.

4. EXPERIMENTAL RESULTS

In this section we first evaluate the proposed interest point detector and point context feature independently in several shape retrieval experiments. After that, we evaluate and compare the performance of the proposed matching method with some traditional point matching methods to illustrate our advantages. In the object retrieval experiments, for each of the shapes used as a query, we have checked whether the retrieved results are correct, i.e., belong to the same class as the query. For quantitative comparison, we kept the experimental convention proposed in [1] and considered the 10 best matches for each query.

4.1 Evaluation of Interest Point Detection

In this subsection, we compare the retrieval performances on Kimia216 [1] dataset using interest points detected by the proposed method and the most related DCE [7] method. Based on the generated interesting points, we compare their retrieval performances using both shape context method [3] and our proposed matching method. Both methods use the same point descriptor and the match algorithm, except the interest point detector. As shown in Table 1, retrieval results based on the proposed interesting point detector perform better in both methods. The main reason is the property of interesting points generated by DCE method is highly related to the stop parameter k (DCE results are collected with $k = 10$ which got the best retrieval result for comparison). Obviously, it is impractical to set appropriate k manually on each object. On the contrary, our method can generate stable interesting points without degree parameters.

SC [3]	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
IP1 [7]	216	210	195	184	181	172	161	146	148	128
IP2	**216**	**212**	**206**	**197**	**191**	**190**	**186**	**186**	**183**	**171**

PC	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
IP1 [7]	216	211	205	196	192	191	186	178	177	175
IP2	**216**	**212**	**211**	**211**	**205**	**200**	**201**	**195**	**193**	**195**

Table 1: Experimental comparison on Kimia216 dataset. Interesting points are detected separately using the proposed method (IP2) and the DCE [7] method (IP1). IP1 and IP2 are then compared using Shape Context (SC) [3] and the proposed Point Context (PC) matching methods.

4.2 Evaluation of Point Context

In order to evaluate the performance of point context descriptor, we compare the retrieval performance using Point Context to other most related descriptors in [3, 5] on MPEG400 dataset. The shapes in this dataset have much larger intra-class variations and inter-class similarities than the MPEG7. Except descriptors, all performances are obtained using the same interest points based on our proposed method and matched by the Hungarian algorithm. As illustrated in Table 2, the point context feature achieves the best performance among three descriptors.

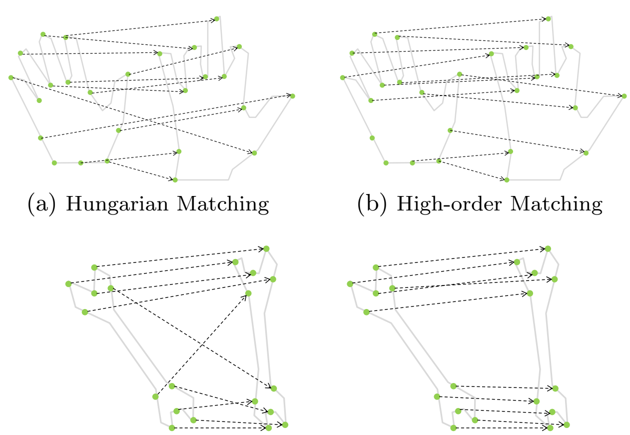
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
SC [3]	370	343	310	302	277	272	265	264	239	240
PCCS [5]	377	351	336	331	317	302	287	282	273	262
Our	**391**	**377**	**372**	**364**	**356**	**343**	**338**	**319**	**304**	**276**

Table 2: Experimental comparison of our point context descriptor to the Shape Context (SC) [3] and partition points-based descriptors (PCCS) [5] using the interesting points generated by our proposed method on MPEG400 dataset. The matching algorithm on all the three methods are Hungarian algorithm.

4.3 Evaluation of High-Order Matching

In this part, with some objects from Kimia99 [2] dataset, we first evaluate the matching performance of the proposed high-order graph matching to the traditional Hungarian matching method. The interesting points in both matching methods are generated and represented by the proposed method. After that, we compare the proposed matching method against the state-of-art method in [5] with the same objects. In Figure 2, we match two hands with some deformations among their fingers. Since there are more interesting points in the left hand than the right one, with our basic constraint in Section 3.1, some points in the left hand will be jumped.

As shown in Figure 2(a), there are some mismatched interesting points because of the similar points in both shapes. Moreover, the geometrical relations among the interesting points are not considered. Figure 2(b) shows that the proposed high-order potentials yield apparently better matching. In Figure 2(c), based on human perception there are several mismatched points. Similar to hands, this is due to their symmetric silhouette and there are also many similar points in one object which could affect the matching performance. In contrast, as shown in Figure 2(d), with the point context feature and our proposed potentials, all points in the left tool can find its correspondences correctly.



(a) Hungarian Matching
(b) High-order Matching
(c) Result in [5]
(d) Our method

Figure 2: Object matching with different potentials and comparing the matching result of the proposed method to [5].

5. CONCLUSION

In this paper, a novel shape matching method based on the point context descriptors and high-order graph matching is presented. The main idea is to generate some interesting points on each shape contour for correspondence matching. The generated interesting points are robust for shape deformation. For each interesting point, we propose a point descriptor, point context, to capture its geometrical locations. In order to involve not only the singleton assignments, but also the geometrical relations of tribes with different interesting points for shape matching, we propose potentials for the singleton and triplet terms.

Acknowledgement. Research activities leading to this work have been supported by the China Scholarship Council (CSC) and the German Research Foundation (DFG) within the Research Training Group 1564 (GRK 1564).

6. REFERENCES

[1] X. Bai and L. Latecki. Path similarity skeleton graph matching. *PAMI*, 30(7):1282–1292, 2008.

[2] X. Bai, L. Latecki, and W. yu Liu. Skeleton pruning by contour partitioning with discrete curve evolution. *PAMI*, 29(3):449–462, 2007.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(4):509–522, 2002.

[4] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce. A tensor-based algorithm for high-order graph matching. *PAMI*, 33(12):2383–2395, Dec 2011.

[5] C. Feinen, C. Yang, O. Tiebe, and M. Grzegorzek. Shape matching using point context and contour segments. In *ACCV*. Springer LNCS, November 2014.

[6] L. Latecki, Q. Wang, S. Koknar-Tezel, and V. Megalooikonomou. Optimal subsequence bijection. In *ICMD*, pages 565–570, Oct 2007.

[7] L. J. Latecki and R. LakÃd'mper. Convexity rule for shape decomposition based on discrete contour evolution. *CVIU*, 73:441–454, 1999.

[8] C. Lu, L. Latecki, N. Adluru, X. Yang, and H. Ling. Shape guided contour grouping with particle filters. In *ICCV*, pages 2288–2295, September 2009.

[9] T. Ma and L. Latecki. From partial shape matching through local deformation to robust global shape similarity for object detection. In *CVPR*, pages 1441–1448, June 2011.

[10] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision*, 40(2):99–121, November 2000.

[11] S. P. Tadeusiewicz R., Ogiela M.R. Notes on a linguistic description as the basis for automatic image understanding. *Int. Journal of Applied Mathematics and Computer Science*, 19(1):143–150, 2009.

[12] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *ECCV*, volume 5303, pages 596–609. Springer Berlin Heidelberg, 2008.

[13] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, 2011.

[14] Y. Zeng, C. Wang, Y. Wang, X. Gu, D. Samaras, and N. Paragios. Dense non-rigid surface registration using high-order graph matching. In *CVPR*, pages 382–389, June 2010.