INVESTIGATIONS ON SKELETON COMPLETENESS FOR SKELETON-BASED SHAPE MATCHING

Cong Yang, Oliver Tiebe, Marcin Grzegorzek*

Institute for Vision and Graphics
University of Siegen
Hoelderlinstr. 3, D-57076 Siegen, Germany

Bipin Indurkhya

Institute of Philosophy
Jagiellonian University
Grodzka 52, 33-332 Krakow,Poland

ABSTRACT

Skeleton is an important shape descriptor for deformable shape matching since it integrates both geometrical and topological features of a shape. As skeletonisation process often generates redundant skeleton branches that may seriously disturb the skeleton matching and cause high computational complexity, skeleton pruning is required to remove the inaccurate or redundant branches while preserving the essential topology of the original skeleton. However, pruning approaches normally require manual intervention to produce visually complete skeletons. As different people may have different perceptions for identifying the visually complete skeletons, it is unclear how much the human selection influences the accuracy of skeleton-based shape matching. Moreover, it is also unclear how skeleton completeness impacts the accuracy of skeleton-based shape matching. We investigate here these two questions in a structured way. In addition, we present experimental evidence to show that it is possible to do automatic skeleton pruning while maintaining the matching accuracy by estimating the approximate pruning power of each shape.

Index Terms— Shape Similarity, Skeleton Matching, Skeleton Pruning, Human Perception

1. INTRODUCTION

Shape representation and matching is a fundamental problem in image processing and computer vision, aaffecting a variety of application domains [1, 2]. However, nearly all of the shape matching approaches face the same challenge: shape deformation. As shown in Figure 1, shapes of the same object are visually different depending on the deformations. To overcome this, many different shape descriptors [3, 4, 5, 6, 7, 8] are proposed to capture both local and global geometric shape properties. Among them, skeleton is an important shape descriptor for deformable shape matching. In addition, shape similarity based on skeleton matching usually performs better than contour or other shape descriptors in the presence of partial occlusion and articulation of parts [9, 10, 11]. The main reason is that skeleton integrates both the geometrical and the topological features of a shape.

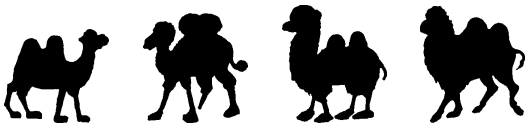
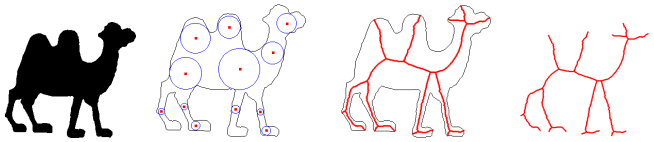


Fig. 1. An illustration of shapes which vary significantly depending on deformations.

A skeleton is defined as a connected set of medial lines along the limbs of its shape [12]. From a technical point of view, such a skeleton is extracted by continuously collecting centre points of maximal tangent disks touching the object boundary in two or more locations, as shown in Figures 2 (b) and (c). The centre point of a maximal tangent disk is referred to as a skeleton point. The sequence of connected skeleton points is called a skeleton branch. A skeleton point having only one adjacent point is a skeleton endpoint. A skeleton point having three or more adjacent points is a junction point. If a skeleton point is not an endpoint or a junction point, it is called a connection point. Based on skeletons, shape similarity can be calculated by matching their skeletons [11, 10, 9]. Specifically, two skeletons are normally matched by considering the topological structures of the skeleton trees or graphs [13, 14, 15, 9].

However, a skeleton is sensitive to the deformation of an object's boundary because a little variation or noise at the boundary often generates redundant skeleton branches that may seriously disturb the topology of the skeleton [16]. Furthermore, a large number of skeleton branches may cause the overfitting problem resulting in a high complexity for skeleton matching [17]. To solve these problems, one approach is to smooth the shape boundary before applying the skeletonisation methods [18]. Though this approach leads to stable skeletons in the presence of boundary deformations, only rough shape matching can be performed since the obtained skeletons do not represent any shape details.

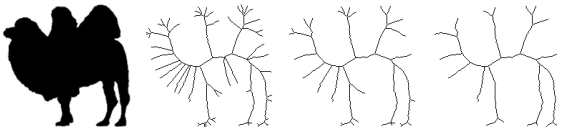
*Research activities leading to this work have been supported by the China Scholarship Council (CSC) and the German Research Foundation (DFG) within the Research Training Group 1564 (GRK 1564).



(a) Shape (b) Maximum disks (c) Center collections (d) Skeleton

Fig. 2. An overview of the skeletonisation process to convert a given shape (a) into a skeleton (d). (b) and (c) visually illustrate the skeleton extraction process.

Another approach is skeleton pruning [19, 16], which removes the inaccurate or redundant branches while preserving the essential topology of the original skeleton. This approach normally requires manual intervention to produce visually pleasing and complete skeletons. For example, in Figure 3, DCE [19] requires a proper stop parameter k to calibrate the pruning power. Different stop parameters for the same object lead to visually different skeletons. In other words, these skeletons have different levels of completeness. In order to generate proper skeleton for shape matching, the stop parameter is selected based on human perception [20, 19]. As different people may have different perceptions for selecting skeletons, it is unclear how much the human selection influences the skeleton-based shape matching. Moreover, it is also unclear how the skeleton completeness impacts the accuracy of skeleton-based shape matching.



(a) Shape (b) $k = 100$ (c) $k = 50$ (d) $k = 25$

Fig. 3. Skeletons on the same camel shape with different DCE [19] stop parameters k.

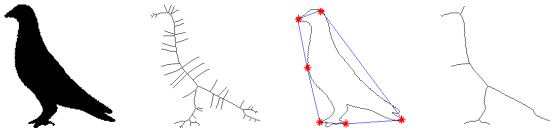
In this paper we study these problems in a structured way: For the first problem, by selecting and voting from different volunteers, perceptually complete skeletons are collected and used for skeleton-based shape matching. For the second problem, given a single shape, we generate several skeletons with different pruning powers. Thus, the generated skeletons have different completeness levels and their matching accuracy can be independently evaluated and compared. Finally, we conclude by comparing the retrieval performances and dissimilarity values in a skeleton-based shape retrieval scenario using the above skeletons.

2. SKELETON PRUNING

We employ the Discrete Curve Evolution (DCE) [19] method for machine-based skeleton pruning, because the pruned skeletons from this method are stable for significant noise and shape variations. For a fair comparison, we also use this method for producing the perceptually complete skeletons with manual intervention.

2.1. Machine-based Skeleton Pruning

Figure 4 gives an overview of the DCE process: (1) Given a planar shape D (Figure 4(a)), the Max-Disk Model [21] is used to generate the initial skeleton $S^n(D)$ (Figure 4(b)) as a set of centre points of circles that are in contact with the shape boundary. That is, $s \in S^n(D)$ is the centre of such a circle, and contact points of s on the shape boundary are the generating points. The first iteration index of DCE is indicated by n, which is iteratively decremented until 3. One of these steps is indicated by k. (2) The boundary of D is regarded as the initial polygon P^n, which will be simplified into a polygon P^k (blue solid line in Figure 4(b)) using the polygon simplification method described below. (3) With P^k, $S^n(D)$ is pruned by removing all skeleton points $s \in S^n(D)$, which contain the generating points in the same contour segment. A contour segment is defined as a part of the shape boundary approximated by the straight line (polygon partition) between two neighbouring convexes of P^k (red stars in Figure 4(c)). Each pruned point s results from a contour segment with respect to the polygon partition, and therefore s can be considered as an unimportant skeleton point and can be removed.



(a) D (b) $S^n(D)$ (c) P^k (d) $S^k(D)$

Fig. 4. Illustration of original shape D, initial skeleton $S^n(D)$, simplified polygon P^k and the pruned skeleton $S^k(D)$ generated by [19] with $k = 10$.

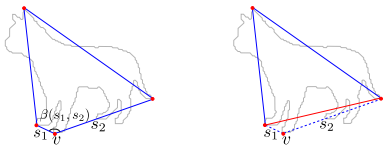


Fig. 5. Polygon simplification. As vertex v has the smallest contribution with Eq. 1, its consecutive line segments s_1, s_2 are replaced by a single line segment (red line).

For the polygon simplification method, the basic idea is as follows: In every step, as shown in Figure 5, a pair of consecutive line segments s_1, s_2 is replaced by a single line segment that connects the endpoints of $s_1 \cup s_2$. Here, v is regarded as having the smallest shape contribution based on the following measure K:

$$K(s_1, s_2) = \frac{\beta(s_1, s_2) l(s_1) l(s_2)}{l(s_1) + l(s_2)} \quad (1)$$

where $\beta(s_1, s_2)$ is the angle of the corner consisting of s_1 and s_2, and l is the length function normalised with respect to the total length of the lines constituting the polygon. Based on K,

The higher the value of $K(s_1, s_2)$, the larger the contribution of $s_1 \cup s_2$ to the polygon. With a shape D, we employ DCE to generate skeletons hierarchically with different k values. (In our experiment, $k \in [3, 14]$).

2.2. Skeleton Pruning with Human Perception

Unlike machine-based pruning, skeleton pruning with human perception not only depends on whether the shapes are represented completely and succinctly, but also considers the shape structural compositions with background knowledge [20]; because the perceptions are easily affected by the surroundings. With these observations, we conduct skeleton pruning with human perception by an individual voting scheme. Specifically, given a shape, based on DCE method, we generate and select three skeletons for voting. According to the observations in [20], the candidates are selected depending on the following constraints: (1) All the intuitive regions in a shape should have at least one skeleton branch. (2) If a shape is globally or regionally symmetric, the skeleton should also be globally or regionally symmetric. (3) In each region, relatively minor contour perturbations should have at most one skeleton branch. (4) Both shape contour and skeleton are clear and visible for the participants.

With the selected candidates, the individual voting scheme is conducted with questionnaires. Specifically, we organise the original shape and the three candidates in a table. In order to fulfil the fourth constraint above, we fuse the skeleton and shape contour together to clarify the original shape structure. We only illustrate the shape number for the subsequent statistics, the shape class and the name are not shown to the participants. For each shape, only one skeleton can be chosen, which is marked by X in the bracket. After statistics, we select the skeleton with the highest votes as the final skeleton from human perception. If two skeletons get the same number of votes, we conduct another round of voting until we get a stable output.

3. SKELETON MATCHING

We employ here the approach proposed in [9] as our skeleton matching method with skeletons from both machine and human perception as explained in Section 2. The basic idea is to find the best matching between the endpoints of two skeletons. Given two skeletons S_1 and S_2, let e_i^1 and e_j^2 be the endpoints in S_1 and S_2, respectively, $i = 1, 2, \cdots, M$, $j = 1, 2, \cdots, N$, $M \geq N$. The skeleton graphs are matched by comparing the geodesic paths [22] between their skeleton endpoints. Then all the dissimilarity costs between their endpoints $c(e_i^1, e_j^2)$ are represented as a distance matrix $M(S_1, S_2)$:

$$M(S_1, S_2) = \begin{pmatrix} c(e_1^1, e_1^2) & c(e_1^1, e_2^2) & \cdots & c(e_1^1, e_N^2) \\ \vdots & \vdots & \vdots & \vdots \\ c(e_M^1, e_1^2) & c(e_M^1, e_2^2) & \cdots & c(e_M^1, e_N^2) \end{pmatrix} \quad (2)$$

The total dissimilarity $c(S_1, S_2)$ between S_1 and S_2 is computed by searching correspondences between the skeleton endpoints with the Hungarian algorithm [23] on $M(S_1, S_2)$ expressed in Equ. 2, so that the endpoints in S_1 and S_2 are matched with the minimal cost. The resulting costs of the matched endpoints can be denoted as $c_1, c_2, \cdots, c_N$ and the total dissimilarity $c(S_1, S_2)$ is calculated by their summation.

4. EXPERIMENT

In this section, we first analyse the availability of the proposed voting method. After that, the skeletons generated by the human perception and machines are employed and compared in a shape retrieval scenario. Finally, the dissimilarity values from shape matchings are compared and analysed.

4.1. Availability of the Individual Voting

Generally, two types of voting schemes can be employed: Individual voting and group voting. The biggest difference between these two schemes is whether the participants can see the votes of others. Specifically, for the group voting scheme, participates are divided into different groups. The participates in each group work on one questionnaire so that the votes are visible for everyone. For our research, we choose the individual voting approach based on the assumption that in the group voting scheme, the articipates' perception could be influenced by the votes from others. In order to verify this assumption, we collected 60 different shapes from 5 classes (glass, camel, elephant, bird, heart) for the experiment. Each class contained 12 different shapes. For the group voting, we divided 35 volunteers into 5 groups. Each group was given 12 randomly selected shapes from different classes using questionnaires. For the individual voting, we used another set of 30 volunteers, each of whom was provided two randomly selected shapes from different classes using the questionnaire introduced in Section 2.2. Based on the statistics, we observe that among 60 shapes, only 38 shapes have the same voting results in the two schemes. The main reason is that in the group voting scheme, if a volunteer sees results from other volunteers, it affects their decision.

4.2. Comparison of Retrieval Results

We apply the skeleton-based shape retrieval scenario using Kimia216 [10] dataset which contains 216 images from 18 classes. Our evaluation is built on a retrieval framework where shapes in the database are ranked based on their similarity to a query shape. To evaluate the retrieval performance, we use the following measure:

$$y = \frac{1}{100} \sum_{n=1}^{Q} R_n (1 - \frac{n-1}{Q}) \quad (3)$$

where Q denotes the number of shapes which belong to the same class as the query shape. R_n denotes the number of retrieved shapes that are in the same class as the query in the top-ranked n shapes. The evaluation measure in Eq. 3 is necessary to evaluate the retrieval performance accurately using both the number of correct matches and ranking positions.

Table 1 depicts the matching performance using the matching method in Section 3 where skeletons are generated by DCE with different stop parameter k and the human perception (the last row). We use each shape as a query and retrieve the 12 most similar shapes among the whole dataset. The final value in each position is the counter value that is obtained by checking the retrieval results using all the 216 shapes as queries. For example, the fourth position in the row of $k = 3$ shows that from 216 retrieval results in this position, 186 shapes are relevant to the query shapes. Scores in the last column are calculated with Eq. 3.

k	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th	Score
3	216	203	196	186	182	154	157	132	117	102	91	77	11.3567
4	216	205	190	182	171	157	148	136	126	117	104	102	11.3208
5	216	203	191	187	184	171	165	154	138	136	117	108	11.7908
6	216	205	188	191	178	181	173	148	148	122	115	98	11.8342
7	216	208	204	197	190	181	156	155	135	127	102	92	12.0067
8	216	205	198	195	197	193	189	176	155	126	120	93	12.3783
9	216	209	199	203	193	191	186	177	162	150	131	118	12.4567
10	**216**	**210**	**209**	**204**	**196**	**197**	**176**	**173**	**164**	**152**	**150**	**110**	**12.5817**
11	216	207	204	196	193	189	177	165	155	141	132	92	12.3550
12	216	209	202	197	188	185	162	163	157	140	137	93	12.2375
13	216	205	206	195	192	186	177	167	152	146	120	89	12.3100
14	216	208	204	199	198	184	185	166	153	158	139	103	12.4917
H	216	203	199	187	180	175	159	153	136	118	113	79	11.7375

Table 1. Experimental comparison of skeleton-based shape matching with different stop parameters k for contour partitioning with DCE. Last row H represents the retrieval results with skeletons generated by human perception.

We can clearly observe that the matching score with skeletons generated by human perception performs only better than skeletons with $k = 3, 4$. This observation tells us that the perceptual skeleton completeness does not have too much influence to the global accuracy of skeleton-based shape matching. In addition, with skeletons generated by machine, the matching score is increasing from $k = 3$ until $k = 10$. After that, the score decreases. The rationale behind this is when the pruning power k is getting smaller, the pruned skeleton is getting less complete and loses more topological and geometrical features of the original shape. On the contrary, when the pruning power k is getting bigger, the pruned skeleton is getting more complete and contains more fine-grained shape features. Thus, the matching accuracy is getting higher. However, if k continues increasing, there are too many endpoints in the skeleton and the overfitting problem appears and impacts the accuracy of skeleton-base shape matching. Therefore, we cannot improve the shape matching accuracy just by simply increasing their skeleton completeness.

4.3. Comparison of Dissimilarity Values

In order to explore the detailed changes within different pruning powers, we calculated the mean dissimilarity values between the objects in each class in Kimia216 dataset [10]. The mean values within the 18 classes are illustrated and compared in Figure 6. Mean dissimilarity values from the skeletons generated by human perception are also compared in Figure 6 (the blue line with H).

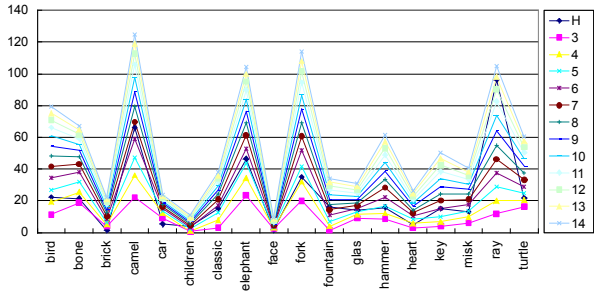


Fig. 6. Comparison of the mean dissimilarity values between skeletons in each class. Skeletons are generated by machine in Section 2.1 with $k \in [3, 14]$ and human perception in Section 2.2.

We can observe that for the machine-generated skeletons, the dissimilarity values increases along with an increase of pruning power, while the changes between different classes remain the same. From this observation we can conclude that for each shape class, the required pruning power is different. In addition, we can also see that the mean values from human perception have almost the same pattern as the other curves in Figure 6. For this reason, if we find a proper range of skeleton pruning power for each class of shape, minor changes of this power do not significantly impact the accuracy of the skeleton-based shape matching.

5. CONCLUSION

We investigated here the influence of skeleton completeness and human perception for skeleton-based shape matching. We compared the shape matching accuracy on Kimia216 dataset between the skeletons from human perception and the machine with different pruning power. Based on the statistics of matching accuracy and dissimilarity scores, we observed that the perceptual skeleton completeness does not have too much influence on the accuracy of skeleton-based shape matching. Moreover, as the skeleton completeness increases, the matching accuracy first increases and then decreases. Finally, we were able to obtain the proper range of skeleton pruning power for each shape class, and observed that minor changes of the pruning power do not effect the overall accuracy of shape matching. Therefore, we can apply the automatic skeleton pruning while maintaining the matching accuracy by estimating the approximate pruning power of each shape.

6. REFERENCES

[1] O. V. Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A survey on shape correspondence," *Computer Graphics Forum*, vol. 30, no. 6, pp. 1681–1707, 2011.

[2] C. Yang, O. Tiebe, P. Pietsch, C. Feinen, U. Kelter, and M. Grzegorzek, "Shape-based object retrieval by contour segment matching," in *IEEE International Conference on Image Processing*, 2014, pp. 2202–2206.

[3] C. Yang, C. Feinen, O. Tiebe, K. Shirahama, and M. Grzegorzek, "Shape-based object matching using point context," in *International Conference on Multimedia Retrieval*, June 2015, pp. 519–522.

[4] J. Xie, P. Heng, and M. Shah, "Shape matching and modeling using skeletal context," *Pattern Recognition*, vol. 41, no. 5, pp. 1756–1767, 2008.

[5] B. Hong and S. Soatto, "Shape matching using multiscale integral invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 151–160, 2015.

[6] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[7] H. Ling and D.W. Jacobs, "Shape classification using the inner-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

[8] Y.L. Zou, C. Li, Z. Boukhers, K. Shirahama, T. Jiang, and M. Grzegorzek, "Environmental microbiological content-based image retrieval system using internal structure histogram," in *International Conference on Computer Recognition Systems*, 2015.

[9] X. Bai and L.J. Latecki, "Path similarity skeleton graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1282–1292, 2008.

[10] T. B. Sebastian, P.N. Klein, and B.B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 550–571, 2004.

[11] T. B. Sebastian and B. B. Kimia, "Curves vs. skeletons in object recognition," *Signal Process*, vol. 85, no. 2, pp. 247–263, 2005.

[12] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities*, Morgan Kaufmann Publishers Inc., 2004.

[13] J. Liu, W. Liu, and C. Wu, "Objects similarity measurement based on skeleton tree descriptor matching," in *Computer-Aided Design and Computer Graphics, 2007 10th IEEE International Conference on*, Oct 2007, pp. 96–101.

[14] P. Dimitrov, C. Phillips, and K. Siddiqi, "Robust and efficient skeletal graphs," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 417–423.

[15] D. Macrini, S. Dickinson, D. Fleet, and K. Siddiqi, "Bone graphs: Medial shape parsing and abstraction," *Computer Vision and Image Understanding*, vol. 115, no. 7, pp. 1044–1061, 2011.

[16] W. Shen, X. Bai, R. Hu, H. Wang, and L. J. Latecki, "Skeleton growing and pruning with bending potential ratio," *Pattern Recognition*, vol. 44, no. 2, pp. 196–209, 2011.

[17] C. Yang, O. Tiebe, M. Grzegorzek, and E. Lukasik, "Skeleton-based audio envelope shape analysis," in *Asian Conference on Pattern Recognition*. 2015, IEEE Computer Society.

[18] D. Geiger, T. Liu, and R.V. Kohn, "Representation and self-similarity of shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 86–99, 2003.

[19] X. Bai, L.J. Latecki, and W. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 449–462, 2007.

[20] C. Firestone and B. J. Scholl, "Please tap the shape, anywhere you like: Shape skeletons in human vision revealed by an exceedingly simple measure," *Psychological Science*, vol. 25, no. 2, pp. 377–386, 2014.

[21] W. Choi, K. Lam, and W. Siu, "Extraction of the euclidean skeleton based on a connectivity criterion," *Pattern Recognition*, vol. 36, no. 3, pp. 721–729, 2003.

[22] L.J. Latecki, Q. Wang, S. Koknar-Tezel, and V. Megalooikonomou, "Optimal subsequence bijection," in *IEEE International Conference on Data Mining*, Oct 2007, pp. 565–570.

[23] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.