Sketch-based 3D Object Retrieval Using Two Views and Visual Part Alignment

Zahraa Yasseen, Anne Verroust-Blondet, Ahmad Nasri

Author manuscript, in Proceedings of the 2015 Eurographics Workshop on 3D Object Retrieval
DOI: 10.2312/3dor.20151053

Sketch–based 3D Object Retrieval Using Two Views and Visual Part Alignment

Z. Yasseen[a,*], A. Verroust-Blondet[a], A. Nasri[b]

[a] *Inria Paris 2 rue Simone Iff CS 42112 75589 Paris Cedex 12 France*
[b] *College of Computing, Fahad Bin Sultan University, KSA.*

Abstract

Hand drawn figures are the imprints of shapes in human's mind. How a human expresses a shape is a consequence of how he or she visualizes it. A query–by–sketch 3D object retrieval application is closely tied to this concept from two aspects. First, describing sketches must involve elements in a figure that matter most to a human. Second, the representative 2D projection of the target 3D objects must be limited to "the canonical views" from a human cognition perspective. We advocate for these two rules by presenting a new approach for sketch–based 3D object retrieval that describes a 2D shape by the visual protruding parts of its silhouette. Furthermore, the proposed approach computes estimations of "part occlusion" and "symmetry" in 2D shapes in a new paradigm for viewpoint selection that represents 3D objects by only the two views corresponding to the minimum value of each.

Keywords: sketch–based 3D object retrieval, 2D Shape description, Best view selection, Symmetry estimation

1. Introduction

The two main components of a sketch–based 3D object retrieval application are the 2D shape description method and the 3D object representation. Despite the extensive variety of shape descriptors proposed in 2D shape retrieval context, a relatively small number of ideas have been exploited in sketch–based 3D object retrieval approaches. The most recurrent 2D shape descriptors are the shape context [3] and the bag–of–features collected from overlapping areas around densely sampled points in the image [15, 8, 7]. Despite the acknowledged advantages of accurate numerical models in general shape definition, a looser abstraction of shapes is needed to deal with entries unregulated by delicate measures.

A part–based 2D shape descriptor introduced in [30] uses the chordal axis transform [22] (CAT) for shape definition and dynamic time warping (DTW) for matching and distance estimation. On an abstract level, the CAT–DTW descriptor starts by a CAT based segmentation of the silhouette of the 2D shapes. The segments or subregions are embedded in a hierarchy to allow a matching–time selection of visual or protruding parts for optimal correspondence. The visual parts are described by geometric attributes and the spatial relations with other parts. CAT–DTW rectifies the semantic gap between shapes of different natures (see Fig. 1) by taking visual part salience measures relative to the constituting shape and its remaining parts.

The matching method of CAT–DTW calculates the distances between visual parts rather than boundary points using a decision dependant DTW technique that rotates the start point to

*Corresponding Author: Zahraa Yasseen; Email, zyasseen@gmail.com

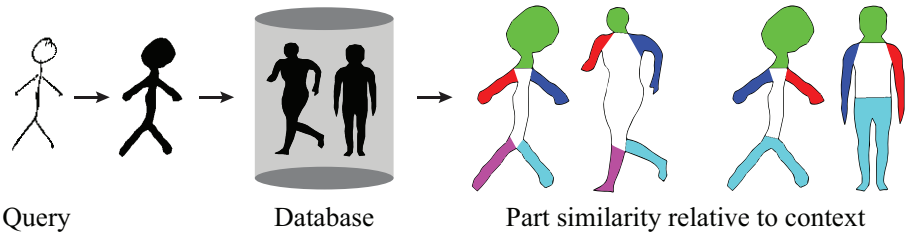


Figure 1: Matching a sketched human stick–figure (after applying erosion and filling) to the silhouettes of 3D models' projections.

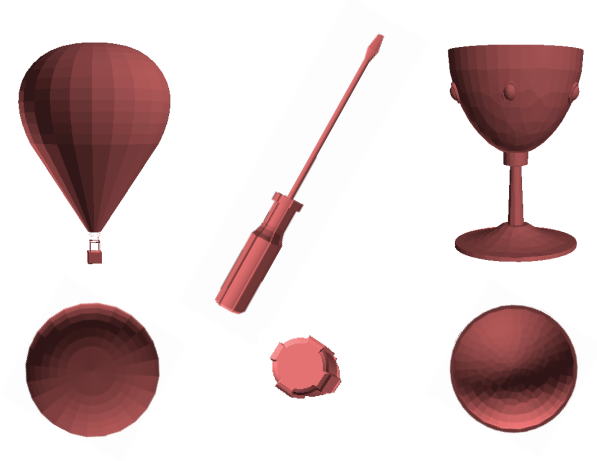


Figure 2: Recognizable (top) and misleading (bottom) views of 3D objects.

find the best match. In this paper, we extend this practice by reversing the direction of search in one of the objects being matched. This reversal allows correct retrievals of similar reflected objects that the original CAT–DTW failed to match. In addition, it facilitates the estimation of the symmetric property of a shape when matched to its inverted version.

Distance between a 2D shape and a 3D object is computed by matching the former to a number of 2D projections representing the latter. State of the art methods have gone as far as retaining one hundred projections and some have even exceeded this number. For a sketch–based 3D object retrieval application, in particular, we seriously question the need for such a number. The excessive number of views not only risks run time efficiency degradation, but also increases the possibility of producing views that mislead the retrieval process. A screwdriver, for example, may and should be represented by one view. A snapshot of such an object taken from an angle along its principal axis deceives even a human inspector (see Fig. 2). A table viewed from the top has a similar shape as a book or a door. It is evident that there are more "adequate" views for a given object but whether these are determined by geometric properties or by learning is yet an open issue. In the light of this discussion, we put forward the necessity to investigate viewpoint selection from human cognition theories' perspective.

Cognitive science approaches the viewpoint selection issue by performing case studies to understand the so called "canonical views". In 1981 Palmer et al. [21] proposed a "maximal information" hypothesis that canonical views are those that give most information about the 3D structure of the object. Blanz et al. [4] experimented with digital 3D models asking the partic-

ipants to rotate and position objects. They concluded that people would try to avoid occlusion of component and seek pronounced asymmetry. The front or side view of symmetric objects such as teapot, cow, or chair rated lowest amongst the selected views. Recently, Mezuman et al. [18] used internet image collections to learn about canonical views and verify precedent theories. Inspired by cognitive science theories, we rely on two concepts to select representative views for a 3D object: minimal part occlusion or maximal information and minimal symmetry. In this paper, we propose methods to quantify these concepts. Taking the projected views from points equidistant to the object's centroid, we relate the level of part occlusion to the sum of lengths of skeletal segments produced by the CAT of each view's silhouette. Symmetry of a given silhouette is estimated by its CAT–DTW distance to its topologically reflected version obtained by a clock–wise (negative direction) traversal of its visual parts.

The rest of the paper proceeds as follows. In section 2, we discuss related work from various aspects. Sections 3 gives an overview of CAT–DTW. A closeup on the details of the DTW technique and the explanation of the topological inversion are presented in section 4. The two–silhouette representation of 3D objects is portrayed in section 5. Evaluation results on SHREC'13 Sketch Track Benchmark and the conclusion follow last.

2. Related Work

The two major subproblems in sketch–based 3D object retrieval are how to obtain the 3D models' 2D representations and what 2D descriptor to use in the matching process. Existing methods may be classified in many ways depending on different approaches adopted to solve subproblems. For 2D data representation, methods either include shapes' internal available details [31, 23, 27, 24, 7, 8] or only analyze the outline [20, 19, 14, 17]. The first class of methods incorporate user strokes inside sketched shapes and include suggestive contours [6], apparent ridges [9], or other computer generated lines in the 2D views of their 3D models. The second class preprocess their 2D data by diluting and filling operations to have one closed contour line and silhouette per 2D sketch or 3D model projection.

Another aspect to classify methods is the dependance on a training stage using the Bag–of–Words model [7, 20, 8]. The opposite class makes direct distance estimation between matched objects using either global [31] or local [27, 7, 20], or both global and local [23, 24, 19, 14, 17, 15] approaches. Global descriptors define a quantization or a feature vector in R^n where the distance metric is defined over that space. Local descriptors represent a shape by a set of feature vectors where the distance is estimated by a minimal cost match between individual features. Methods that use both global and local employ the global descriptor in a pruning stage.

View selection of 3D models has also been tackled in different ways. In general, two motivations have guided this process. The first is to include as many views as feasible so as not to miss a potential viewing angle selected by a human user to draw the object. These methods either select corners and edge midpoints on the bounding box [31, 23, 27, 24] or generate uniformly sampled points on the bounding sphere [20, 14, 8] with viewing direction pointing towards the center. The second motivation is to find views more likely to be used by humans and reduce the number of generated images. Napoleon et al. [19] first align the model and then take only up to 9 projections. Eitz et al. [7] employ Support Vector Machine with a radial basis function kernels to learn a "best view classifier" during the training stage and use it in the testing stage. Li et al. [14] use the View Context similarity between the sketch and saved projections to prune unlikely views in an alignment stage. In a later publication, and following the observation that not all 3D models views are equally important, Li et al. [17] propose a complexity metric based on viewpoint entropy distribution. The idea is to assign more views for more complex objects and thus recommend class–specific numbers of projections.

A recent family of methods has emerged characterized by employing machine learning methods to bridge the semantic gap between sketches and projection images. Li et al. [15] use a Support Vector Machine with radial basis function kernel to build a classifier that predicts the possibilities of the input sketch belonging to all the categories. Furuya et al. [8] use a semi–supervised machine learning method called Manifold Ranking Algorithm [32]. The algorithm works by diffusing relevance value from the query to the 3D models in a Cross–Domain Manifold the two domains being sketches and 3D models.

Since year 2012, sketch–based 3D shape retrieval contests (SHREC) are being held on yearly basis [13, 11, 16, 12]. A participating group would contribute in more than one run showing results of different parameter settings or choice of particular algorithms. It is notable that there is a small range of 2D shape descriptors tested in sketch–based 3D object retrieval compared to the much larger number of available choices. The 2D shape descriptor that we employ in this paper uses a skeleton to represent shapes by visual parts and their spacial relationships.

3. 2D Shape Description

In order to keep this paper as self contained as possible, we give an overview of the CAT–based shape description method. However, more details can be found in the original CAT–DTW documentation [30].

The input data is a binary image representing the silhouette of a single object. We extract the contour, locate corner points, and sample the in–between contour fragments uniformly. The advantage of locating corner points is the inclusion of the sharp features in the sample set. The region is then triangulated using Constrained Delaunay Triangulation (CDT). The rectified CAT and a set of pruning and merging operations provide a skeleton with an association between skeletal segments and subregions (see Fig. 3(a)). Subregions are categorized according to their connectivity into three types: terminal, sleeve, and junction characterized by one, two, and many connected segments respectively.

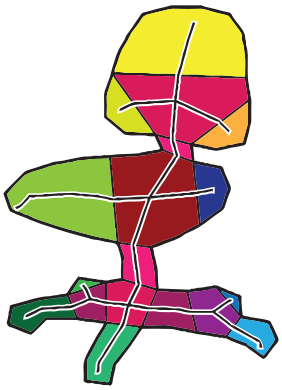
The CAT segments are embedded in a tree where leaf nodes correspond to terminal segments (see Fig. 3(b)). We leave out the process that locates the root of the tree since it does not influence the course of this paper. Our main concern is the visual parts of the shape and how they are represented in this hierarchy. First, terminal nodes with relative size, eccentricity, and convexity beyond some thresholds are labeled as salient nodes. Starting from the bottom of the tree, the visual parts of the shape are represented by all subtrees that constitute less than two salient nodes. Visual parts that contain more than one node in their subtrees represent a set of CAT segments joined into one higher level entity denoted by a *wing* node (see Fig. 3).

The visual parts, comprised of terminal and wing nodes that we denote by feature nodes, are kept in their anti–clockwise order of appearance along the boundary of the object. Each node is described by 9 geometric attributes: area, perimeter, eccentricity, circularity, rectangularity, convexity, solidity, bending energy, and chord length ratio in addition to a radial distance signature. These values are combined into a feature vector v that is made of two parts: geometric parameters p and the radial distance signature r. The distance between two vectors v_1 and v_2 is the Euclidian distance between the parameters plus the squared distance between the signature part.

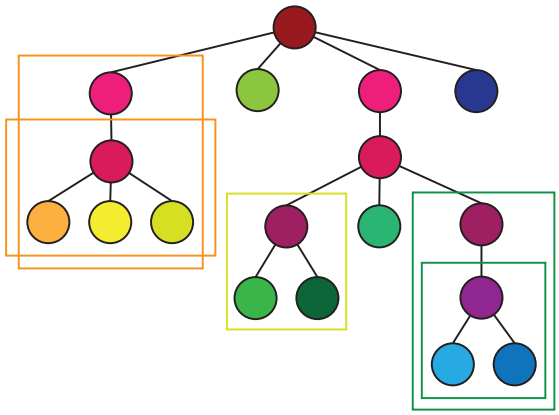
$$d(v_1, v_2) = sqrt\left[\sum_{i=1}^{9}(p_1[i] - p_2[i])^2\right] + \sum_{i=1}^{15}(r_1[i] - r_2[i])^2 \quad (1)$$

Similarly, the norm of the feature vector is given by:

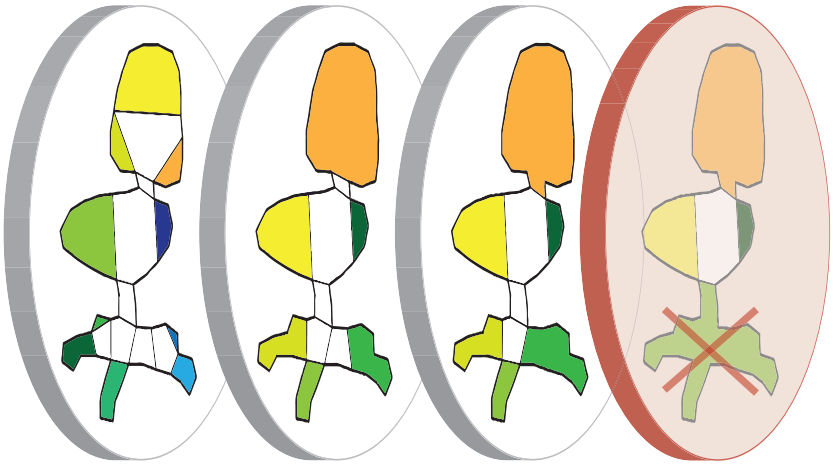
$$|v| = sqrt\left[\sum_{i=1}^{9} p[i]^2\right] + \sum_{i=1}^{15} r[i]^2 \quad (2)$$



(a) The CAT and the subsequent segmentation.



(b) Marked subtrees correspond to wing nodes. The leaves are arranged from left to right in the anti–clockwise order of appearance along the boundary of the shape.



(c) Visual parts represented by terminal nodes on the finest level of detail (left) and by wing nodes on higher levels. More than two salient nodes cannot be included in the same wing (right) and thus stop the process of wing node formation.

Figure 3: The visual parts embedded in a hierarchical structure. The tree nodes in Fig. 3(b) are shaded with the same color of their corresponding subregion in Fig. 3(a).

The spatial and angular distances between feature nodes comprise an inter–distance matrix relating every pair of them. An entry (i, j) in this matrix is a 3 dimensional vector (d_E, d_{BE}, dA) corresponding to Euclidian distance, bending energy, and angular distance between nodes i and j.

4. Adapted Dynamic Time Warping Method

Dynamic time warping is a method that originated in the context of aligning voice signals with different time latency [28]. Later on, it was introduced to the shape matching world to measure distance between closed shapes. Roughly, the idea is to rotate one shape while calculating a distance matrix for every obtained alignment. Each row corresponds to the distance between a point in the first shape and all points in the other. A minimal distance path is calculated for every matrix resulting in a point–to–point or point–to–segment pairing. The matrix that produces the minimal distance among others represents the best alignment.

The feature nodes are represented in a feature space of dimension N ($N = 24$) comprised of an assembly of geometric parameters. Every object has an ordered set of feature vectors in addition to an inter–distance matrix. To match two shapes A and B, we seeks a set of pairs associating feature nodes from A and B. The couples correspond to non–overlapping visual parts and must not violate their anti–clockwise ordering. The cost of a match is defined by the sum of the following values:

1. The distance defined in Eq. 1 between coupled feature nodes from A and B.
2. For every consecutive couple $(q_{(q\in A)}, r_{(r\in B)})$ and $(s_{(s\in A)}, t_{(t\in B)})$, the internal distance defined by: $(d_E(q, s) - d_E(r, t))^2 + (d_{BE}(q, s) - d_{BE}(r, t))^2 + (d_A(q, s) - d_A(r, t))^2$
3. Terminal nodes that are not included in the match cost a penalty equal to the norm in Eq. 2.

Every terminal node in each object is a potential starting point for the anti–clockwise traversal of feature nodes. To find the optimal solution, we compute the minimal cost matches for all possible combinations of starting terminal nodes of the two shapes. However, due to their relation with wing nodes, some terminal nodes are excluded from the set of candidate start points. In the following sections, we describe what viable configurations are and how the cost matrix is built and handled.

4.1. Generating Viable Configurations

Wing nodes are visual features that must be considered for matching as a whole in any tested configuration. A terminal node selection as the starting point should not cause any of its related wing nodes to be split between the beginning and the end of the list of feature nodes. This observation leads to the introduction of the *stop point* which is a terminal that has either one of the following properties:

- It does not belong to any wing node.
- It is the first terminal node to appear in the anti–clockwise direction in all the wings it belongs to.

Different configurations are generated by alternately shifting one object's start node to the next stop point while fixing the other.

4.2. Decision–based Dynamic Time Warping

Every configuration provides two ordered sequences of feature nodes to be matched. The dynamic time warping technique finds the minimal cost path by setting up a matrix of all possible matches. Starting from (n, m) towards $(0, 0)$, the cost of the optimal path is accumulated following the minimal cost path rule defined by: $cost(i, j) = cost(i, j) + min(cost(i + 1, j + 1), cost(i + 1, j), cost(i, j + 1))$ Our variation of the solution follows from the specifics of the problem.

We construct an $n \times m$ matrix where n and m are the numbers of terminal nodes of the two shapes. Every entry in this matrix contains a decision node that enumerates all possible options that can be taken when the entry is reached. The decision node compares the cost of a terminal–terminal, terminal–wing, wing–terminal, wing–wing, and a void match. The void match is the decision to exclude one or both of the terminals from the matching process. This list of options is not independent from its surrounding matrix entries. For example, a wing–wing matching decision affects the matrix block spanned by the terminals constituting these two wings (see Fig. 4). This slightly alters the minimal cost path rule since at (i, j), the "previous" entry is not simply either one of $(i + 1, j + 1)$, $(i + 1, j)$, or $(i, j + 1)$. It is rather related to the option at hand and the block of matrix spanned by the nodes being matched according to this option. After all decision nodes have selected their minimal cost option, the optimal cost of the current configuration is found in the minimal cost at entry $(0, 0)$.

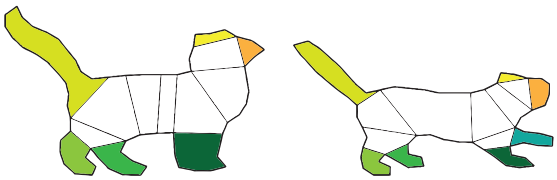
4.3. Topological Inversion

As described so far, CAT–DTW works well on Kimia–99 and Kimia–216 2D shape datasets [26]. However, it happens that these datasets do not include reflected instances of the same class. For example, the correct match between the shapes shown in Fig. 5 will never be found using the current CAT–DTW. The visual parts of these two objects are arranged in reversed orders: head, tail, hind legs, front legs for the first object and head, front legs, hind legs, tail for the second. Reversing the direction of terminal traversal for one of the objects allows obtaining the configuration that would give the optimal match as shown in Fig. 5. When an object is matched to its inverted version, the distance is an estimate of the degree of symmetry. Smaller values indicate stronger symmetric property of the shape (some examples are shown in Fig. 6). We call the inverse of this distance *the symmetry measure* and use it to find asymmetric projected views of 3D objects as shown in the next section.

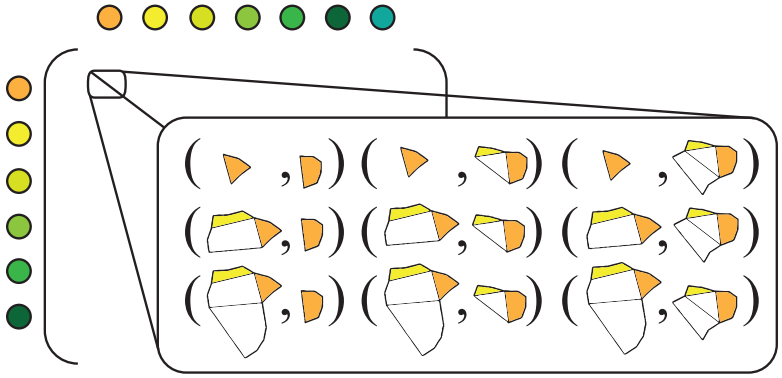
5. 3D Object's Representative Views

We take projected images of the 3D object from 50 views positioned on the unit sphere bounding the object and pointing towards its center. First, the object is scaled and translated to lie within a cube half the size of the unit cube. Then one Catmull–Clark subdivision [5] step is applied to the cube producing a volume defined by 26 vertices and 24 faces. The vertices and the centroids of the faces are translated in the radial direction so that they all lie on the unit sphere and equidistant from the origin. Each viewpoint gives a binary silhouette representation of the 3D object.

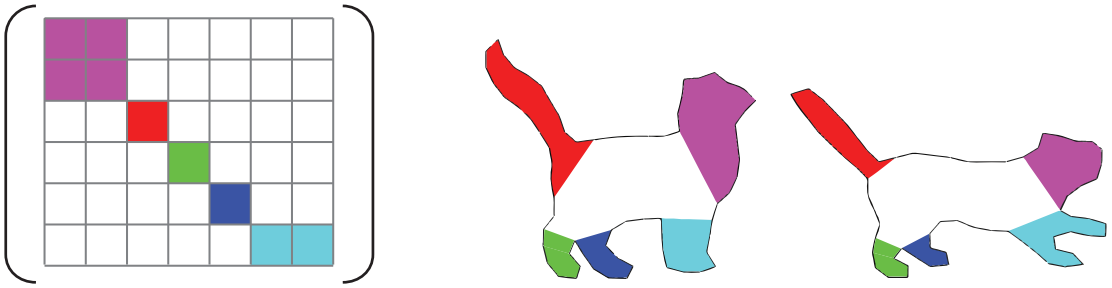
When humans design sketches to represent an object, they tend to make all the meaningful salient parts of the object visible (the four legs of the cow/horse, the legs of a chair, etc.) even if the perspective view of the object is altered. This is a demonstration of the "minimal part occlusion" theory proposed as one of the "canonical views" criteria. Following this observation, we select the silhouettes having the greater skeletal length which we compute as the sum of skeletal segments of terminal and sleeve nodes and the maximal three skeletal segments of junction nodes. The silhouettes with top k skeletal lengths (k equal to 10 in our experiments)



(a) Configuration where the start points are the snouts in both cats.



(b) The decision matrix where a row (respectively column) corresponds to a terminal in the first (respectively second) object. Each entry (i, j) holds all possible pairings between the visual parts related to the terminals associated to row i and column j.



(c) The minimum cost path in the matrix and the consequent part correspondence between the matched shapes.

Figure 4: The optimal correspondence between two shapes obtained from the minimal cost path in the distance matrix. Note how the 9^{th} option at entry $(0, 0)$ shown in Fig. 4(b) gives a minimal cost and leads to the pairing highlighted in purple in Fig. 4(c).

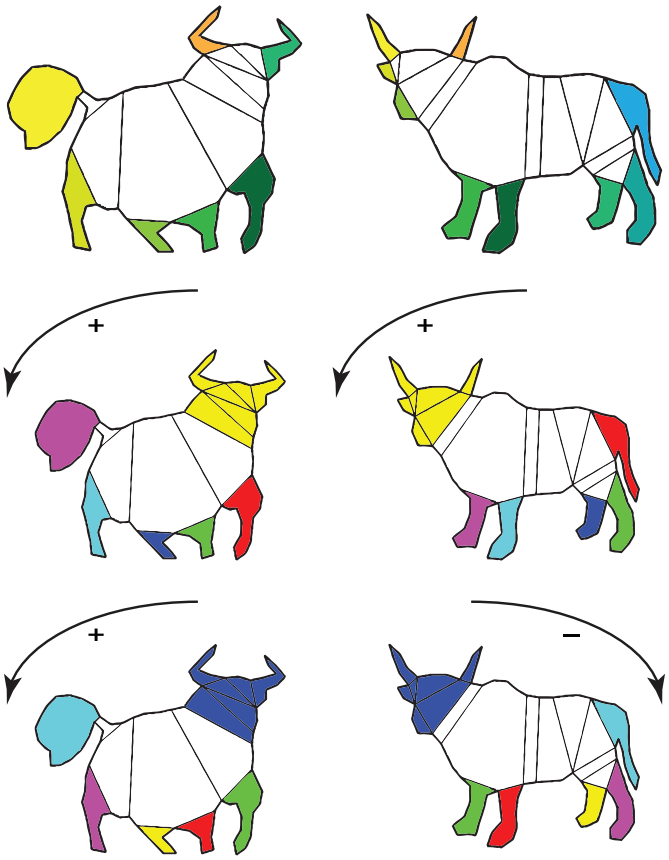


Figure 5: Applying DTW to find part correspondence between the objects shown top row where the visual parts' orderings are *horn, tail, hind leg, ..., horn* and *horn, horn, front leg, ..., tail* respectively. The method matches the *heads* correctly due to rotating the start point of the second object so as to have the two *horns* adjacent. However, due to reflectance, all other visual parts are mismatched. The third row shows the setting where the second object is arranged in the reverse direction. The total distance obtained in this setting is minimal and the visual parts are paired more accurately.

Figure 6: Symmetric shapes and their associated asymmetry evaluation. Lower values indicate stronger symmetry.

PARTICIPANT	METHOD	NN	FT	ST	E	DCG	AP
Aono, Masaki [2]	EFSD	0.023	0.019	0.036	0.019	0.24	0.031
Li, Bo [14]	SBR-2D-3D-NUM-50	0.132	0.077	0.124	0.074	0.327	0.094
Li, Bo [17]	SBR-VC-NUM-100	0.164	0.097	0.149	0.085	0.348	0.113
	SBR-VC-NUM-50	0.132	0.082	0.131	0.075	0.331	0.098
Saavedra, Jose M. [24]	FDC	0.053	0.038	0.068	0.041	0.279	0.051
Furuya [8]	BF-fGALIF	0.176	0.101	0.156	0.091	0.354	0.119
	BF-fGALIF+BF-fDSIFT	0.213	0.123	0.186	0.107	0.379	0.143
	CDMR-BF-fGALIF	0.242	0.174	0.263	0.146	0.427	0.215
	CDMR-BF-fGALIF+CDMR-BF-fDSIFT	0.279	0.203	0.296	0.166	0.458	0.246
	UMR-BF-fGALIF	0.159	0.119	0.179	0.102	0.367	0.131
	UMR-BF-fGALIF+UMR-BF-fDSIFT	0.209	0.131	0.195	0.113	0.386	0.152
Our method	CAT-DTW	0.220	0.122	0.180	0.101	0.379	0.128

Table 1: Performance metrics for the performance comparison on the testing dataset of the SHREC'13 Sketch Track Benchmark.

are selected into a candidate set S_k and the rest are discarded (see Fig. 7). Two silhouettes remain to be selected from S_k such that the first is the one with maximal skeletal length and the second has minimal symmetry measure.

6. Experimental Results and Discussion

We tested the 2D shape descriptor and our view selection paradigm on the testing datasets of the SHREC'13 Sketch Track Benchmark [11]. Our 2D shape descriptor handles closed shapes with no holes. For both the sketches and 3D models' projections, we perform filling operations to produce a single contour for analysis. Moreover, we apply a series of erosion and filling operations on sketches to amend disconnected boundary lines and give more emphasis to strokes expressing thin features such as tails or antennas (see Fig. 8). When surrounding entities are sufficiently disconnected from the main depicted object (see the barn in Fig. 8), they are discarded by taking the extracted boundary line that has the greatest length. This works well with this sketch dataset since it happens that in such cases, secondary entities are drawn smaller than the main object.

We employ the seven performance metrics adopted in SHREC'13 [11]. They are Precision–Recall (PR) diagram, Nearest Neighbor (NN), First Tier (FT), Second Tier (ST), E–Measures (E), Discounted Cumulated Gain (DCG) and Average Precision (AP). To compute these metrics, we use the evaluation code available from the contest's website. Table 1 shows that our approach outperforms the methods tested on the same benchmark except for those that employed machine learning by cross–domain manifiold ranking (CDMR). However, the average response time per query of our method is 27.79 seconds on an Intel Core i7 3632QM @ 2.20GHz 8GB RAM while the CDMR employing methods exceed 600 seconds on an Intel Core i7 3930K @ 3.20GHz 64GB RAM. In addition, the precision recall plot in Fig. 9 shows that our method performs best amongst its peers.

Compared to other methods that participated in this track, Saavedra et al [24] use the least number of sample views for a 3D model. They use the 6 orthogonal views (top, bottom, left,

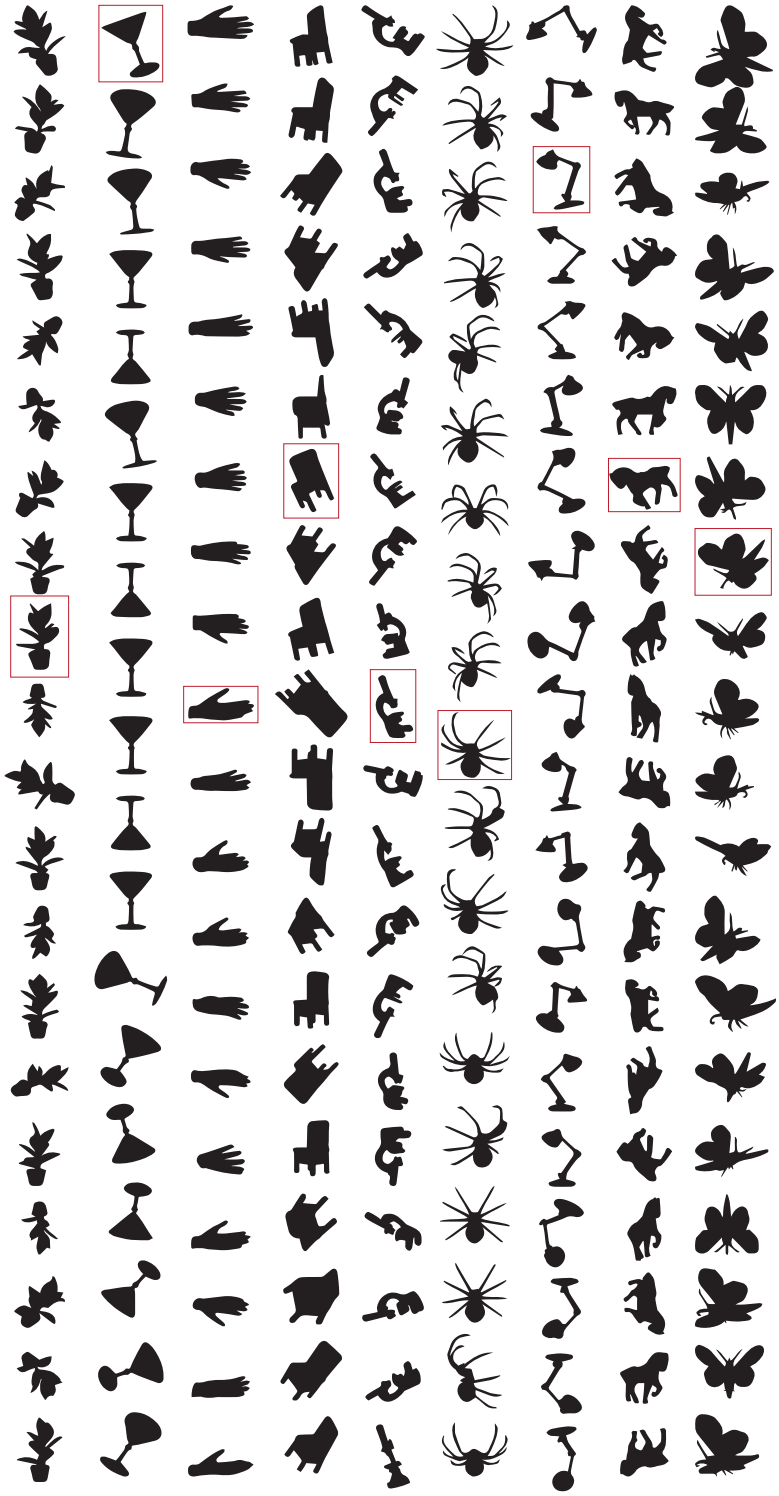


Figure 7: The silhouettes with the top 20 skeletal lengths. For each object, the representative views are the one with maximal skeletal length (first silhouette in the column) and the silhouette with minimal symmetry selected from the top 10 skeletal lengths (marked by the red box).

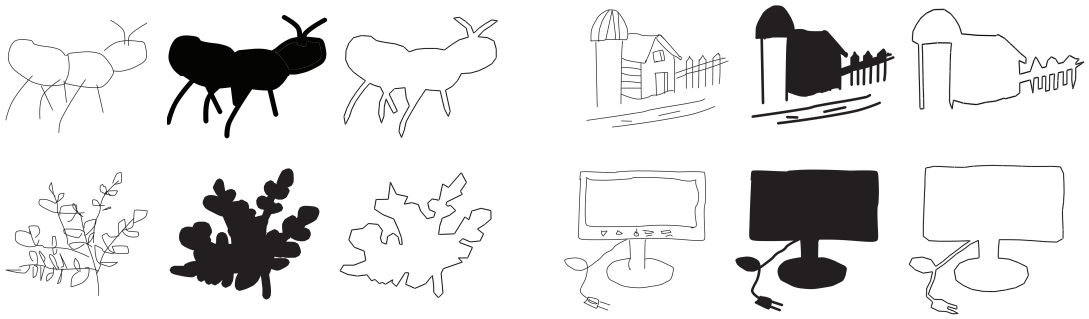


Figure 8: Contour line extraction of sketched images.

right, front, and back). However, their method's performance evaluation reveals the shortcomings of this choice. It is evident that without a suitable alignment method, the orthogonal views of a 3D object cannot give any guarantees that they include a *canonical* view as visualized, and consequently depicted, by humans. Despite increasing the number of views to 26, Aono et al. [2] still score lowest on the precision recall plot diagram. On the other hand, Li et al. [14] (SBR-2D-3D-NUM-50) start form 81 sample views for each 3D object and attempt to align each to the query sketch retaining the best 4 candidates. In another method (SBR-VC-NUM) [17], they drop the alignment stage and keep a precomputed number of sample views per class. The performance improvement of this method (SBR-2D-3D-NUM-50 to SBR-VC-NUM-50) is negligible. Furuya et al. [8] use the highest number of views proposed in this field (162 views) and still need machine learning to improve their retrieval results increasing the retrieval time in an enourmous leap (0.49 seconds for BF-fGALIF to 615.95 seconds for CDMR-BF-fGALIF).

Reporting better performance over these methods while using only two sample views, we verify the merit of the "informative" and "asymmetric" criteria in viewpoint selection. In addition, two other hypothesis are supported by these results. The first one is the logical opposite of more views implying better performance. On the contrary, there are *incorrect* views for 3D models that cause misinterpretation and mismatching and thus *must* be eliminated from its set of sample views. The second hypothesis is the propriety of a visual part–based shape descriptor for a query–by–sketch retrieval of 3D objects. This does not draw from the performance metrics alone but rather from the fact that this descriptor behaves poorly with classes characterized by weak part salience. Nonetheless, it still managed to compensate this setback and produce overall better results.

7. Conclusion

We proposed a sketch–based 3D object retrieval approach that outperforms the methods that contributed in SHREC'13 [11] on the testing dataset of the Sketch Track Benchmark. We showed that a descriptor based on salient parts, their relative sizes and protrusion angles is essential to match conceptually similar but precisely dissimilar objects, which is the case with sketch–based retrieval applications. In addition, we demonstrated that an excess in 2D representations of 3D objects has a potentially degrading effect on the performance results of any method. We made intra–object matches between its projections and composed criteria based on the notion of *informative and asymmetric views* to represent the object by only two views.

The system at hand is liable to many improvements subject to further experiments. Throughout its successive stages other methods for sampling, segmentation, shape signatures, and part

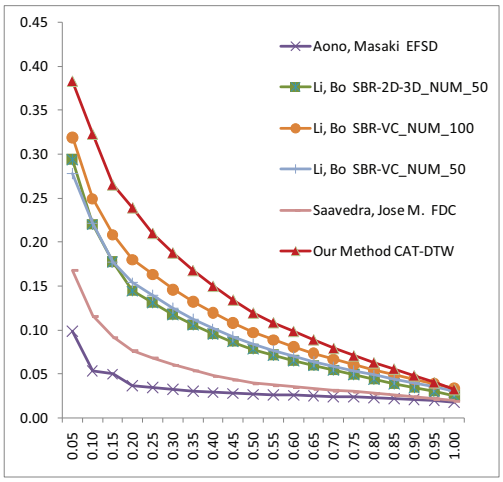


Figure 9: Precision–Recall diagram performance comparisons on the testing datasets of the SHREC'13 Sketch Track Benchmark.

correspondence can be tested. The devised algorithm generates all possible configurations and search for the optimal match of each. Many methods for complexity reduction have been proposed in the general framework of DTW [1, 29, 25, 10]. In addition to these methods, some pruning strategies can be applied to avoid the detailed correspondence computations for each configuration.

References

[1] Ghazi Al-Naymat, Sanjay Chawla, and Javid Taheri. Sparsedtw: A novel approach to speed up dynamic time warping. In *Proceedings of the Eighth Australasian Data Mining Conference - Volume 101*, AusDM '09, pages 117–127, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc.

[2] Masaki Aono and Hiroki Iwabuchi. 3d shape retrieval from a 2d image as query. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC) 2012*, volume 3, 2012.

[3] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NIPS*, volume 2, page 3, 2000.

[4] Volker Blanz, Michael J Tarr, Heinrich H Bülthoff, and Thomas Vetter. What object attributes determine canonical views? *Perception-London*, 28(5):575–600, 1999.

[5] Edwin Catmull and James Clark. Recursively generated b-spline surfaces on arbitrary topological meshes. *Computer-aided design*, 10(6):350–355, 1978.

[6] Doug DeCarlo, Adam Finkelstein, Szymon Rusinkiewicz, and Anthony Santella. Suggestive contours for conveying shape. *ACM Trans. Graph.*, 22(3):848–855, July 2003.

[7] Mathias Eitz, Ronald Richter, Tamy Boubekeur, Kristian Hildebrand, and Marc Alexa. Sketch-based shape retrieval. *ACM Trans. Graph.*, 31(4):31:1–31:10, July 2012.

[8] Takahiko Furuya and Ryutarou Ohbuchi. Ranking on cross-domain manifold for sketch-based 3d model retrieval. In *CW*, pages 274–281. IEEE, 2013.

[9] Tilke Judd, Frédo Durand, and Edward Adelson. Apparent ridges for line drawing. *ACM Trans. Graph.*, 26(3), July 2007.

[10] Daniel Lemire. Faster retrieval with a two-pass dynamic-time-warping lower bound. *Pattern Recogn.*, 42(9):2169–2180, September 2009.

[11] B. Li, Y. Lu, A. Godil, T. Schreck, M. Aono, H. Johan, J. M. Saavedra, and S. Tashiro. Shrec'13 track: Large scale sketch-based 3d shape retrieval. In *Proceedings of the Sixth Eurographics Workshop on 3D Object Retrieval*, 3DOR '13, pages 89–96, Aire-la-Ville, Switzerland, Switzerland, 2013. Eurographics Association.

[12] B Li, Y Lu, C Li, A Godil, T Schreck, M Aono, M Burtscher, H Fu, T Furuya, H Johan, et al. Extended large scale sketch-based 3d shape retrieval. In *Eurographics Workshop on 3D Object Retrieval*, pages 121–130. The Eurographics Association, 2014.

[13] B. Li, Tobias Schreck, Afzal Godil, Marc Alexa, Tamy Boubekeur, Benjamin Bustos, J. Chen, Mathias Eitz, Takahiko Furuya, Kristian Hildebrand, S. Huang, H. Johan, Arjan Kuijper, Ryutarou Ohbuchi, Ronald Richter, Jose M. Saavedra, Maximilian Scherer, Tomohiro Yanagimachi, G. J. Yoon, and Sang Min Yoon. Shrec'12 track: Sketch-based 3d shape retrieval. In *3DOR*, pages 109–118, 2012.

[14] Bo Li and Henry Johan. Sketch-based 3d model retrieval by incorporating 2d-3d alignment. *Multimedia Tools and Applications*, 61(1), November 2012. online first version.

[15] Bo Li, Yijuan Lu, and Ribel Fares. Semantic sketch-based 3d model retrieval. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, pages 1–4. IEEE, 2013.

[16] Bo Li, Yijuan Lu, Afzal Godil, Tobias Schreck, Benjamin Bustos, Alfredo Ferreira, Takahiko Furuya, Manuel J. Fonseca, Henry Johan, Takahiro Matsuda, Ryutarou Ohbuchi, Pedro B. Pascoal, and Jose M. Saavedra. A comparison of methods for sketch-based 3d shape retrieval. *Computer Vision and Image Understanding*, 119(0):57 – 80, 2014.

[17] Bo Li, Yijuan Lu, and Henry Johan. Sketch-based 3d model retrieval by viewpoint entropy-based adaptive view clustering. In *Proceedings of the Sixth Eurographics Workshop on 3D Object Retrieval*, 3DOR '13, pages 49–56, Aire-la-Ville, Switzerland, Switzerland, 2013. Eurographics Association.

[18] Elad Mezuman and Yair Weiss. Learning about canonical views from internet image collections. In *Advances in Neural Information Processing Systems*, pages 719–727, 2012.

[19] Thibault Napoléon and Hichem Sahbi. From 2d silhouettes to 3d object retrieval: contributions and benchmarking. *J. Image Video Process.*, 2010:1:1–1:22, January 2010.

[20] Ryutarou Ohbuchi and Takahiko Furuya. Scale-weighted dense bag of visual features for 3d model retrieval from a partial view 3d model. *In IEEE ICCV 2009 workshop on Search in 3D and Video (S3DV)*, pages 63 –70, 2009.

[21] S. Palmer, E. Rosch, and P. Chase. Canonical perspective and the perception of objects. *Attention and performance IX*, pages 135–151, 1981.

[22] Lakshman Prasad. Rectification of the chordal axis transform skeleton and criteria for shape decomposition. *Image and Vision Computing*, 25(10):1557 – 1571, 2007. Discrete Geometry for Computer Imagery 2005.

[23] Jose Saavedra, Benjamin Bustos, Maximilian Scherer, and Tobias Schreck. Stela: sketch-based 3d model retrieval using a structure-based local approach. In *Proc. ACM International Conference on Multimedia Retrieval (ICMR'11)*, pages 26:1–26:8. ACM, 2011.

[24] Jose M. Saavedra, Benjamin Bustos, Tobias Schreck, Sang Min Yoon, and Maximilian Scherer. Sketch-based 3d model retrieval using keyshapes for global and local representation. In *3DOR*, pages 47–50, 2012.

[25] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11(5):561–580, October 2007.

[26] Thomas B. Sebastian, Philip N. Klein, and Benjamin B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):550–571, May 2004.

[27] Tianjia Shao, Weiwei Xu, KangKang Yin, Jingdong Wang, Kun Zhou, and Baining Guo. Discriminative sketch-based 3d model retrieval via robust shape matching. *Comput. Graph. Forum*, 30(7):2011–2020, 2011.

[28] TK Vintsyuk. Speech discrimination by dynamic programming. *Cybernetics and Systems Analysis*, 4(1):52–57, 1968.

[29] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Min. Knowl. Discov.*, 26(2):275–309, March 2013.

[30] Z. Yasseen, A. Verroust-Blondet, and A. Nasri. Shape matching by part alignment using extended chordal axis transform. *Pattern Recognition*, 57:115 – 135, 2016.

[31] Sang Min Yoon, Maximilian Scherer, Tobias Schreck, and Arjan Kuijper. Sketch-based 3d model retrieval using diffusion tensor fields of suggestive contours. In *Proceedings of the international conference on Multimedia*, MM '10, pages 193–200, New York, NY, USA, 2010. ACM.

[32] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. *Advances in neural information processing systems*, 16(16):321–328, 2004.