View selection for sketch-based 3D model retrieval using visual part shape description

Zahraa Yasseen, Anne Verroust-Blondet, Ahmad Nasri

Author manuscript, The final publication is available at Springer via
http://dx.doi.org/10.1007/s00371-016-1328-7

View Selection for Sketch–based 3D Model Retrieval Using Visual Part Shape Description*

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] *National Council for Scientific Research, CNRS — 59, Zahia Salmane street, Jnah — P.O. Box 11-8281, Beirut, Lebanon.*

Abstract

Hand drawings 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 should 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, we present a list of candidate 2D projections that represent the canonical views of a 3D object. The general rule is that humans would visually avoid part occlusion and symmetry. We quantify the extent of part occlusion of the projected silhouettes of 3D objects by skeletal length computations. Sorting the projected views in the decreasing order of skeletal lengths gives access to a subset of best representative views. We experimentally show how views that cause misinterpretation and mismatching can be detected according to the part occlusion criteria. We also propose criteria for locating side, off axis, or asymmetric views.

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

1. Introduction

3D object retrieval is an evolving domain motivated by the need to manage rapidly growing repositories of 3D models. The basic idea is to find a feature space in which a 3D object has a numeric representation. This allows quantifying the similarity between different 3D objects. The similarity or distance measure, termed as a shape descriptor, may produce a classification over the dataset and facilitate its indexing. However, in order to fetch a 3D model, another similar 3D model has to be provided as a query. This is called querying by example and is considered impractical when it comes to retrieval tasks. Using keywords is not the solution. Most 3D models are saved under irrelevant filenames. Thus came the need for a sketch-based retrieval system that uses hand drawn sketches as queries to find 3D objects similar to the depicted shape.

When a three dimensional object is projected to the 2D world, information still trivial to the human eye is lost in the process. Examples are the z-depth, the viewing angle, and the occluded parts. Hand-drawn sketches escalate difficulties by inaccuracy and imprecision. Classical shape

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

analysis techniques such as corner detection and curvature computations become unreliable. Another added difficulty is the absence of a unified way of expression. When different subjects draw a sketch of the same object, the outcome is not evident. If the object is a "human", some would draw a stick figure while others might depict a more informative figure. One might represent a "cat" by its entire body or else, only its head. Even if the types of drawings are similar, the angle of perspective remains an issue. Which angle would one choose if asked to draw a "chair"?

The two main components of a sketch–based 3D object retrieval application are the 2D shape description and the method that produces 2D projections of the 3D objects. The shape descriptor must have the capacity to overcome the semantic gap between precise computer generated images and erroneous sketches to perform matching. The majority of existing techniques use histograms in the shape descriptor. They perform exhaustive information extraction to capture shape properties. Despite the acknowledged advantages of using local features, internal relativity is needed to deal with the semantic diversity between matched objects. The overall concept is to decompose the shape into parts that are defined in their context relative to the shape and other parts. There is also a necessity to preserve the shape's topology in this part decomposition. In our previous work [38], we show how shape matching by part correspondence allows an extent of variability between the natures of matched objects. We employ a part–based 2D shape descriptor introduced in [37] that uses a chordal axis transform method [28] (CAT) to define and dynamic time warping (DTW) to match. On an abstract level, the CAT-DTW descriptor starts by a skeleton 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 spatial relations with other parts and ordered according to their anti–clockwise appearance along the boundary. 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. Man made sketches are represented best by a descriptor that gives more significance to the elements that matter most to humans. This argument is supported in [38] by outperforming histogram methods by a retrieval approach that describes shapes by salient features while using only two views for 3D shape representation.

The customary practice for the 2D representations of 3D objects is to select viewing angles around an object and take snapshots as representative views. The number of snapshots used [18, 16, 17] per 3D object ranges from 7 through 102 and up to 162. 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 snapshot of a screwdriver, for example, 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. The importance of a viewpoint is assessed by the amount of information it reveals. It gradually varies from a "a good representation" to "misleading". In this paper, we experimentally show how views with minimal information about the object lead to a general performance degradation. We argue that for a sketch-based 3D object retrieval application, in particular, the general practice of uniformly distributing viewing points around a 3D object is diluting useful information in the abundance. Had it been an image matching problem, our claim would not have held owing to the fact that images might have been taken from unknown angles. However, in this case, it is a human subject searching for a 3D object by depicting the image that this subject visualizes in her/his mind.

Cognitive science approaches the viewpoint selection issue by performing case studies to understand the so called "canonical views". In 1981 Palmer et al. [27] 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

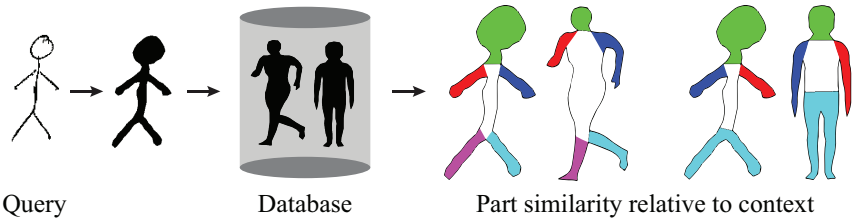


Figure 1: Matching a sketched human stick–figure (after applying erosion and filling) to the silhouettes of 3D models' projections.

participants 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. Mezuman et al. [23] used internet image collections to learn about canonical views and verify precedent theories. Recently, Zhao et al. [40] attempted to learn best views from sketches by asking subjects to align a 3D model according to a given sketch.

Inspired by cognitive science theories, we relied in [38] on two concepts to select representative views for a 3D object: minimal part occlusion or maximal information and minimal symmetry. We quantify the level of part occlusion by the sum of lengths of skeletal segments of each view's silhouette. Symmetry of a given silhouette is estimated by its distance to its topologically reflected version obtained by a clock–wise (negative direction) traversal of its visual parts. In this paper, we perform a full study on the set of views using the testing datasets of the SHREC'13 Sketch Track Benchmark [16]. First, we take projected views from 50 points equidistant to the object's centroid and sort them in their decreasing skeletal lengths (SL) order. We then experiment with the number of views representing an object and evaluate using performance metrics. The general conclusion that we reached is that up to the sixth greatest SL views performance increases rapidly. Then the slope of performance decreases gradually until a climax is reached between 25 and 30 views. Afterwards, all performance metrics start a decreasing slope reaching a minimum when we use all 50 views to represent a 3D object.

This paper is an extended study of what was originally proposed in [38] where we elaborate more on the shape descriptor introduced in [37] and augmented in [38], and highlight some setbacks in the query dataset. Furthermore, we present a thorough analysis of the 2D projected silhouettes of 3D objects, and study the influence of the number of representative views and the impact of the view selection criteria on the retrieval performances. In addition, we present the results of applying our method to the extended large scale dataset used in SHREC'14 [17]. The rest of the paper proceeds as follows. In Section 2, we discuss related work from various aspects. Section 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. A diagnostic description on the testing datasets and the experimental results on the 2D representations of 3D objects are reported and analyzed in Section 5. Comparison with other methods applied on SHREC'13 and SHREC'14 Sketch Track Benchmarks and the conclusion follow last.

2. Related Work

Shape definition is a problem that has been addressed and tackled in different fields of research such as cognitive science, image classification, and 3D object retrieval. In the first field, the requirement is a mathematical representation of shapes that is as close as can be to human's cognition. Experimental validation is carried out by statistics on human subjects' observations

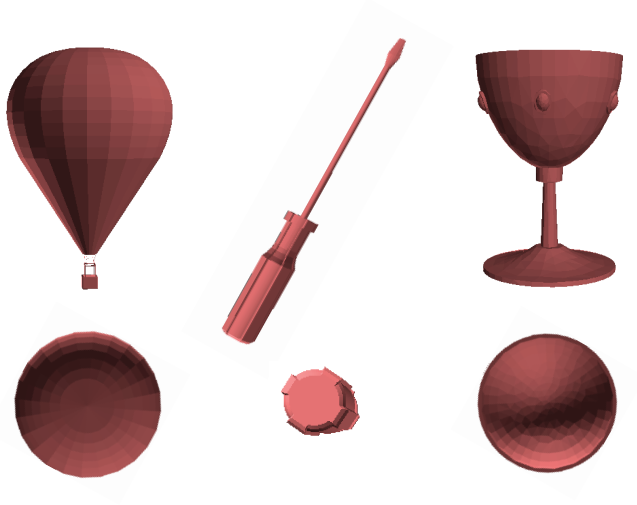


Figure 2: Recognizable (top) and misleading (bottom) views of 3D objects.

regarding collections of stimuli. There are many studies [13, 7, 8, 25, 3] supporting the theory of human's shape perception by parts and the role of skeletons in shape's segmentation. The other two fields are more closely related. However, despite the wide variety of proposed 2D shape descriptors and the reported need for major advances in 3D object retrieval [17], a relatively small number of ideas have been exploited. Reported performance results on existing sketch–based object retrieval approaches have revealed a need for a sketch interpretation component. A question is raised on what existing 2D shape descriptors can offer in terms of understanding human ways of depicting shapes. In this section, we review sketch–based 3D object retrieval approaches with a detailed close up on their shape description mechanisms.

The two major subproblems in sketch–based 3D object retrieval are:

- How to obtain the 2D representations of the 3D model?
- What 2D descriptor to use in the matching process?

Existing methods may be classified in many ways depending on different approaches adopted to solve these subproblems. For 2D data representation, existing methods either include shapes' internal available details [39, 29, 33, 30, 11, 12] or only analyze its outline [26, 24, 19, 22, 42]. The first class of methods incorporates user strokes inside sketched shapes and includes suggestive contours [9], apparent ridges [14], or other computer generated lines in the 2D views of their 3D models. The second class of these methods preprocesses their 2D data by diluting and filling operations to get one closed contour line and silhouette per 2D sketch or 3D model projection.

Another aspect to classify previous methods is the dependance on a training stage using the Bag–of–Words model [11, 26, 12, 42]. The opposite class makes direct distance estimation between matched objects using either global [39] or local [33, 11, 26], or both global and local [29, 30, 24, 19, 22, 20] 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 [39, 29, 33, 30] or generate uniformly sampled points on the bounding sphere [26, 19, 12, 42] with viewing direction pointing towards the center. The second motivation is to distinguish the viewing points that would be more preferred by humans and limit the generated projections to these views. Napoleon et al. [24] first align the model and then take only up to 9 projections. Eitz et al. [11] 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. [19] use the View Context similarity between an input sketch and the 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. [22] 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. Zou et al. [42] reduce the number of sample views by minimizing a representation error of a basic view. They note that the number of views they obtain per class are in accordance with those in [22]. Recently, Zhao et al. [40] devise a supervised learning method to automatically select an average of 7 views of 3D shapes.

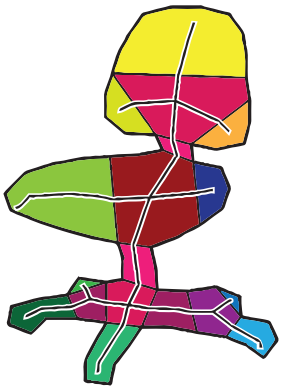
A recent family of methods has emerged characterized by employing machine learning methods to bridge the semantic gap between sketches and projection images. Li et al. [20] use a Support Vector Machine with radial basis function kernel to build a classifier that predicts the possibilities of the input sketch belonging to all the categories. Furuya et al. [12] use a semi–supervised machine learning method called Manifold Ranking Algorithm [41]. The algorithm works by diffusing relevance value from the query to the 3D models in a Cross–Domain Manifold where the two domains are sketches and 3D models. Tatsuma and Aono participated in SHREC'14 [17] employing an extended manifold ranking method (SCMR) that improves the recall. For each 3D object they compute depth buffer images from 102 viewpoints and test their Overlaped Pyramid of Histograms of Orientation Gradients (OPHOG) method on the sketch images alone.

Since year 2012, sketch–based 3D shape retrieval contests (SHREC) are being held on yearly basis [18, 16, 21, 17]. 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 [37].

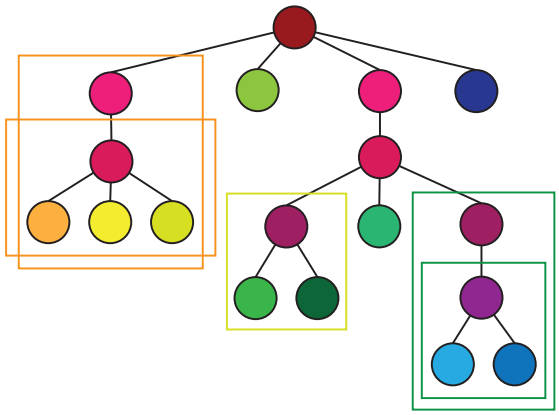
3. 2D Shape Description

For the sake of completeness, we give an overview of the CAT–based shape description method. However, a detailed description of our 2D shape matching method CAT-DTW and its performance evaluation on Kimia-99 and Kimia-216 2D shape datasets [32] can be found in [37].

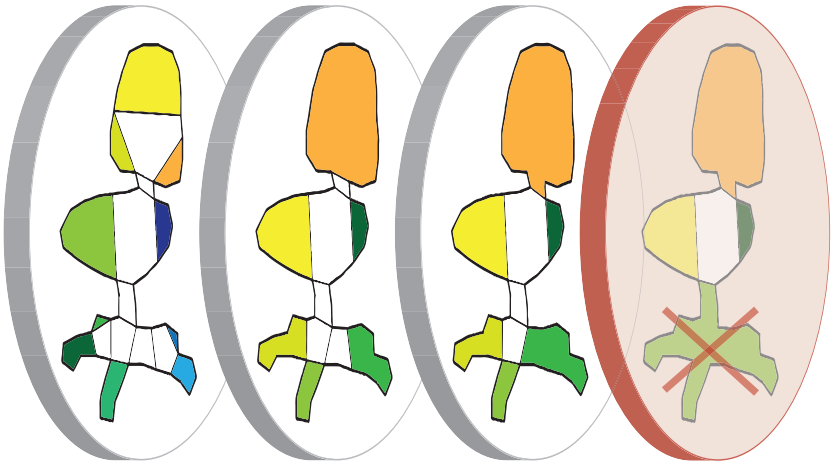
The input data is a binary image representing the silhouette of a single object. We extract the contour, locate corner points, and sample the in–between contour fragments uniformly. The advantage of locating corner points is the inclusion of the sharp features in the sample set. The region is then triangulated 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.



(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).

3.1. Visual Parts

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. The levels of detail may differ substantially between matched shapes. For example, a 3D object of a human figure may be modeled with its intricate details including fingers. A scribbled sketch, on the other hand, barley contains strokes depicting the arms. In the silhouette of the object with fine details, fingers are features and must be included in the visual parts set. However, during the matching process fingers do not have corresponding parts in the query sketch. Accordingly, the process must search for a topologically more complex match for the arm in the sketch. Hence came the necessity of including subtrees in addition to the leaves in the visual parts set of the shape. Nonetheless, not all subtrees are eligible to enter the matching process. Therefore, subtrees are selected into the visual parts set as follows. 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).

3.2. Numeric Representation

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.

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

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

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

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 respectively.

A shape is thus described by a set of vectors representing its visual parts in addition to a matrix of inter–distances between parts. In the following section, we describe how we use the dynamic time warping method to align the visual parts of two shapes. This alignment yields a distance estimated by the cost of the optimal correspondence between their parts.

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 [35]. 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 seek a set of pairs associating feature nodes from A and B. Our problem definition for matching two objects is as follows. Let two shapes A and B with their ordered set of feature nodes each denoted by:

$$F^A = \{f_i^A,\ i = 1 : n\}$$

and

$$F^B = \{f_j^B,\ j = 1 : m\}$$

An alignment between A and B is expressed by a set of couples (f_i^A, f_j^B) subject to the following constraints:

1. Feature nodes are completely disjoint: f_i^A and f_{i+1}^A must not overlap.
2. An alignment is an ordered set of pairs such that f_i^A precedes f_{i+1}^A in the anti-clock wise direction and similarly, f_i^B precedes f_{i+1}^B.

An alignment does not necessarily include all features in both shapes. It is possible to have a best alignment between two objects where some parts are completely left out of the correspondence couples. The cost of a match is defined by the sum of the following values:

1. The distance defined in Eq. 1 between f_i^A and f_j^B.
2. The internal distance defined by:

$$\begin{aligned}(d_E(f_i^A, f_{i+1}^A) - d_E(f_j^B, f_{j+1}^B))^2\ \ +\\ (d_{BE}(f_i^A, f_{i+1}^A) - d_{BE}(f_j^B, f_{j+1}^B))^2\ \ +\\ (d_A(f_i^A, f_{i+1}^A) - d_A(f_j^B, f_{j+1}^B))^2\end{aligned}$$

3. A penalty equal to the norm in Eq. 2 for each terminal node that is not included in the match.

To find the optimal solution, we compute the minimal cost for all possible alignments. Trivially, every terminal node in each object is a potential starting point for the anti–clockwise traversal of feature nodes. 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 for each alignment 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.

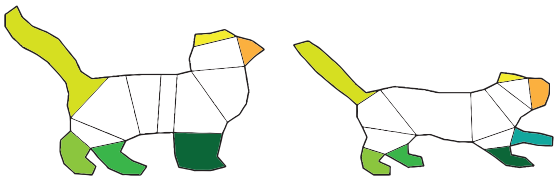
We construct an $n \times m$ matrix where n and m are the numbers of terminal nodes of the two shapes. Every entry in this matrix contains a decision node that enumerates all possible options that can be taken when the entry is reached. The decision node compares the cost of a terminal–terminal, terminal–wing, 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. Symmetry Measure

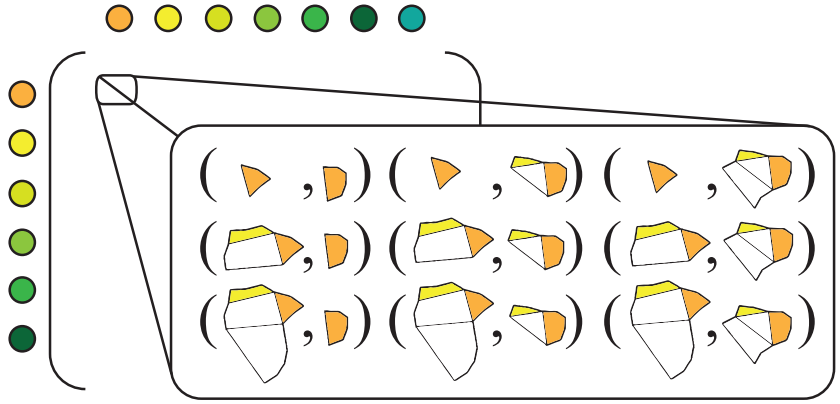
As described so far, CAT–DTW works well on Kimia–99 and Kimia–216 2D shape datasets [32]. However, it happens that these datasets do not include reflected instances of the same class. For example, the correct match between the shapes shown in Fig. 5 will never be found using the current CAT–DTW. The 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

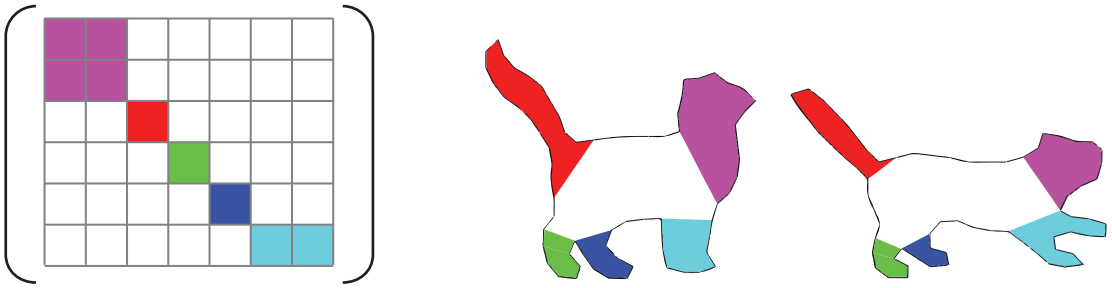
When humans mentally align an object, they tend to make all the meaningful salient parts of the object visible (the four legs of the cow/horse, the legs of a chair, etc.) even if the perspective view of the object is altered. This is a demonstration of the "minimal part occlusion" theory proposed as one of the "canonical views" criteria. We quantify this property by the skeletal length (SL) which we compute as the sum of skeletal segments of terminal and sleeve regions and the maximal three skeletal segments of junction regions (see Fig. 8). Since these silhouettes are taken from equidistant points around the center of the object, there is no need to re-scale their SL values.



(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).

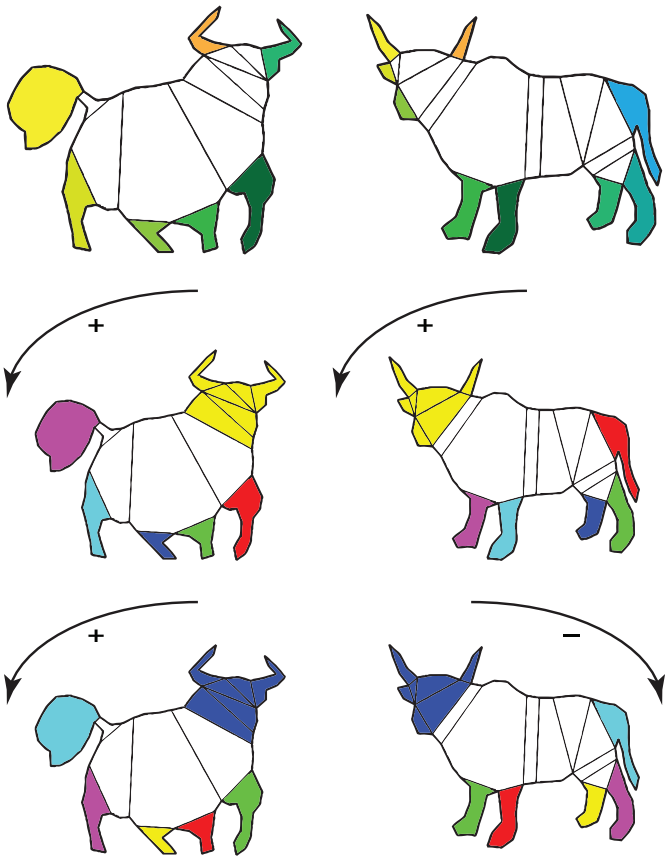


Figure 5: Applying DTW to find part correspondence between the objects shown top row where the visual parts' orderings are *horn, tail, hind leg, ..., horn* and *horn, horn, front leg, ..., tail* respectively. The method matches the *heads* correctly due to rotating the start point of the second object 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.

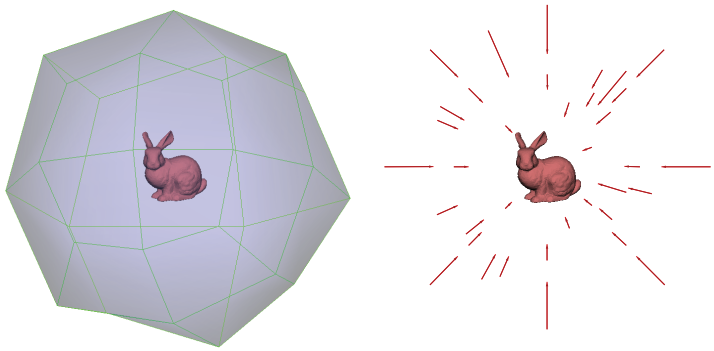


Figure 7: Cameras positioned on a bounding volume directed towards object's center.

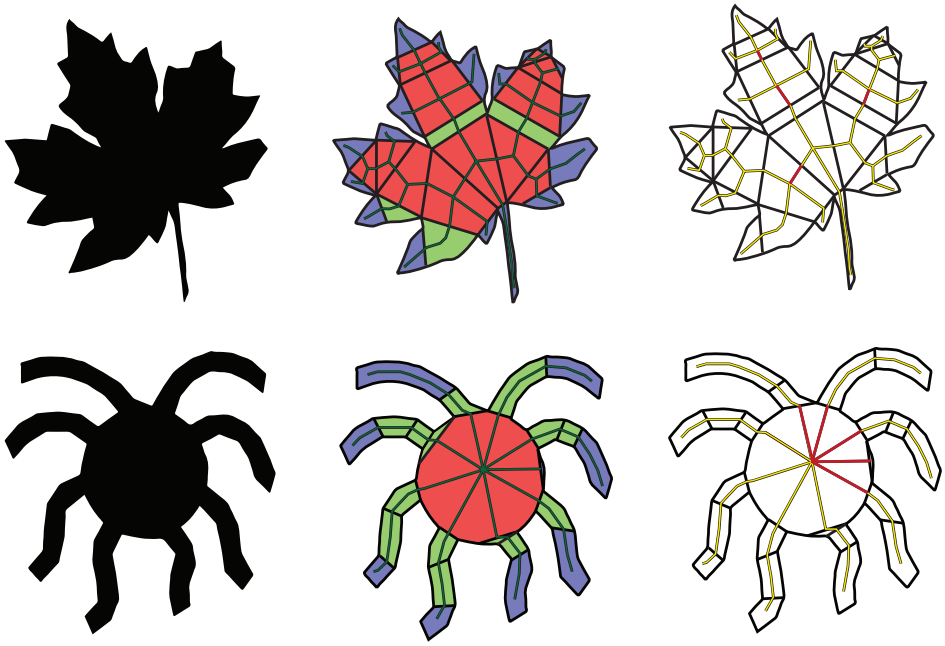


Figure 8: The CAT (skeleton) and the consequent segmentation of two shapes with different topologies. The terminal, sleeve, and junction regions are in blue, green, and red respectively. The SL values for each shape is computed as the sum of lengths of the skeletal segments in yellow. The segments in red are excluded from the computations.

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 (see Fig. 7). 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.

The 50 silhouettes are sorted in the decreasing order of their SL values. Fig. 12 shows the first 20 and the last six of the ordered silhouettes of 14 objects. The particularly significant silhouettes are the first and the last ones of each. The silhouette with maximal SL is the one that defies part occlusion most. This property diminishes gradually towards the end of the list. The silhouette with minimal SL corresponds to the viewing angle which is closest to be aligned along the object's axis. The applications of specifying this angle are beyond the scope of this paper. However, we give some examples shortly to verify its potentials.

5.1. Testing Datasets

We tested the 2D shape descriptor and our view selection paradigm on the testing dataset used in SHREC'13 [16] built on the Princeton Shape Benchmark dataset [34] and comprising 1258 selected models distributed on 90 classes.

Hand sketched figures used in 3D object retrieval experiments are line drawings that are sometimes reduced to unfaithful scribbles. These sketches were collected by Eitz et al. [10] in an attempt to study how humans sketch objects and to build a classifier that performs well with them. Although the subjects that drew these figures were given instructions not to add context around depicted objects and to produce recognizable figures, for many cases these instructions were not followed. In addition, the authors performed a perceptual study and reported that humans correctly identified the object category of a sketch 73% of the time. This means that there is a considerable amount (about 27%) of sketches that cannot be recognized by a human eye. Fig. 9 shows a sample of unrecognizable sketches that we picked from this dataset. Furthermore, there are numerous cases where context is strongly present in the drawing (see Fig. 10). Dragons blow fire, space shuttles have trailing flames, trains are drawn with railroads or smoke, and submarines with surroundings like fish, bubbles, or water level. The rate of these occurrences is relatively high reaching 50% for the dragon, space shuttle, and train classes. It is evident that a sketch drawn for the sole purpose of retrieving a 3D object would not have these additions as opposed to retrieving an image where context is strongly effective.

Despite its impeding properties, the largest yet sketch dataset assembled in [10] has been considered the reference testing benchmark for 3D object retrieval methods. The 3D model dataset used in SHREC'13 [16] was built on the Princeton Shape Benchmark dataset [34] from which 1,258 models were matched to 90 classes out of the 250 classes of the sketch dataset [10]. An extended large scale dataset was collected in SHREC'14 [17] to match 171 classes with the total of 8,987 3D models.

We experiment on the testing dataset of SHREC'13 [16] to study the effect of the number of views representing a 3D object. The testing dataset is comprised of 2,700 sketches distributed uniformly over the same 90 classes of the target 1,258 3D models. We also report the results of the evaluation on the complete dataset of SHREC'14 Sketch Track Benchmark consisting of 13,680 sketches and 8,987 3D models.

5.2. Silhouette Extraction

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.

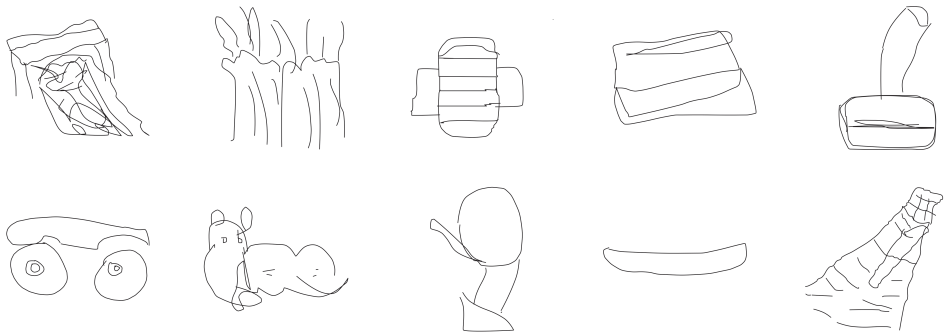


Figure 9: A sample of unrecognizable sketches: bridge, bush, cabinet, couch, microscope, pickup truck, rabbit, satellite dish, skateboard, skyscraper.

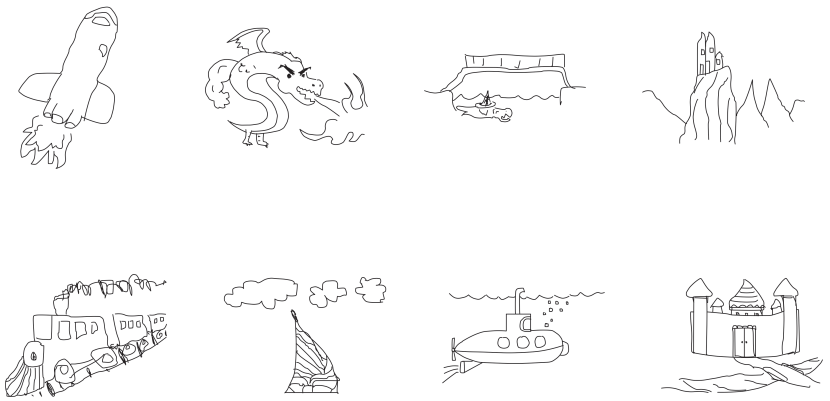


Figure 10: A sample of sketches where context is strongly present.

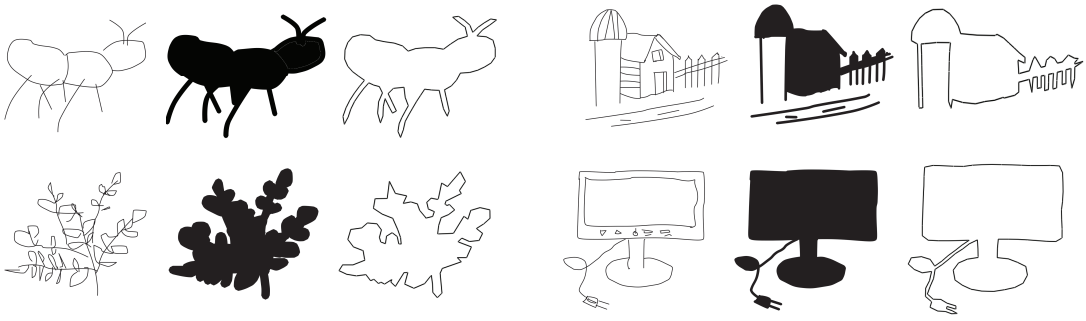


Figure 11: Contour line extraction of sketched images. First, apply a series of erosion and filling operations to obtain closed silhouettes. Then extract boundaries and retain the longest as the single representative contour line.

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. 11). This process reduces identifiability of shapes of the following types:

- sketches drawn within context (sketches in Fig. 10) since the silhouette will include all attached surroundings.
- 3D objects and sketches with low distinctive topology (book or door) and defined by internal feature lines or strokes.
- 3D objects with higher genus and defined by number and size of holes (ladder and beer mug). However, for sketches, boundary lines cannot be distinguished from feature lines indicating ridges and valleys. So, holes may be deemed useless when it comes to sketch–based retrievals.

When surrounding entities are sufficiently disconnected from the main depicted object (see the barn in Fig. 11), 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.

6. Experimental Results and Analysis

The 50 views of each 3D object are sorted in the decreasing order of SL. We performed a series of 13 tests by starting with only the first view with maximal SL and then progressively adding more views. In each setting, we use the 2700 sketches in the testing dataset of SHREC'13 [16] as queries and employ the seven performance metrics adopted in the same contest [16]. 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). We expand the representative views set of a 3D object to include the second, third, sixth, and tenth greatest SL views recomputing the evaluation metrics for every experiment. We continue the tests adding 5 views at a time until all fifty views are used to represent a 3D object in the target dataset.

The graphs shown in Fig. 13 reveal weak improvement for view numbers greater than 10. We compute the NN of every class for the view numbers less than 10. Fig. 14 shows the number of views that obtained the highest NN values. 61% of the classes performed best with view numbers less than 3. Furthermore, the precision recall plot (see Fig. 15) for views number 3 shows considerable improvement over views numbers 1 and 2. Whereas the gain obtained over 3 views by taking the 6 views and the 10 views representations is insignificant. Doubling and tripling the target dataset size (from 3 to 6 views and 10 view respectively) produces more time and data size overhead than performance gain. The small range of difference between 10 views and 25 views on one side and 10 views and 3 views on the other side offer the following observations:

- A 3D object can be represented efficiently by only 3 views.
- The top 10 informative views form a rich subset to select representative views from.

The benefit of these observations will appear shortly as we investigate different criteria for the object representation.

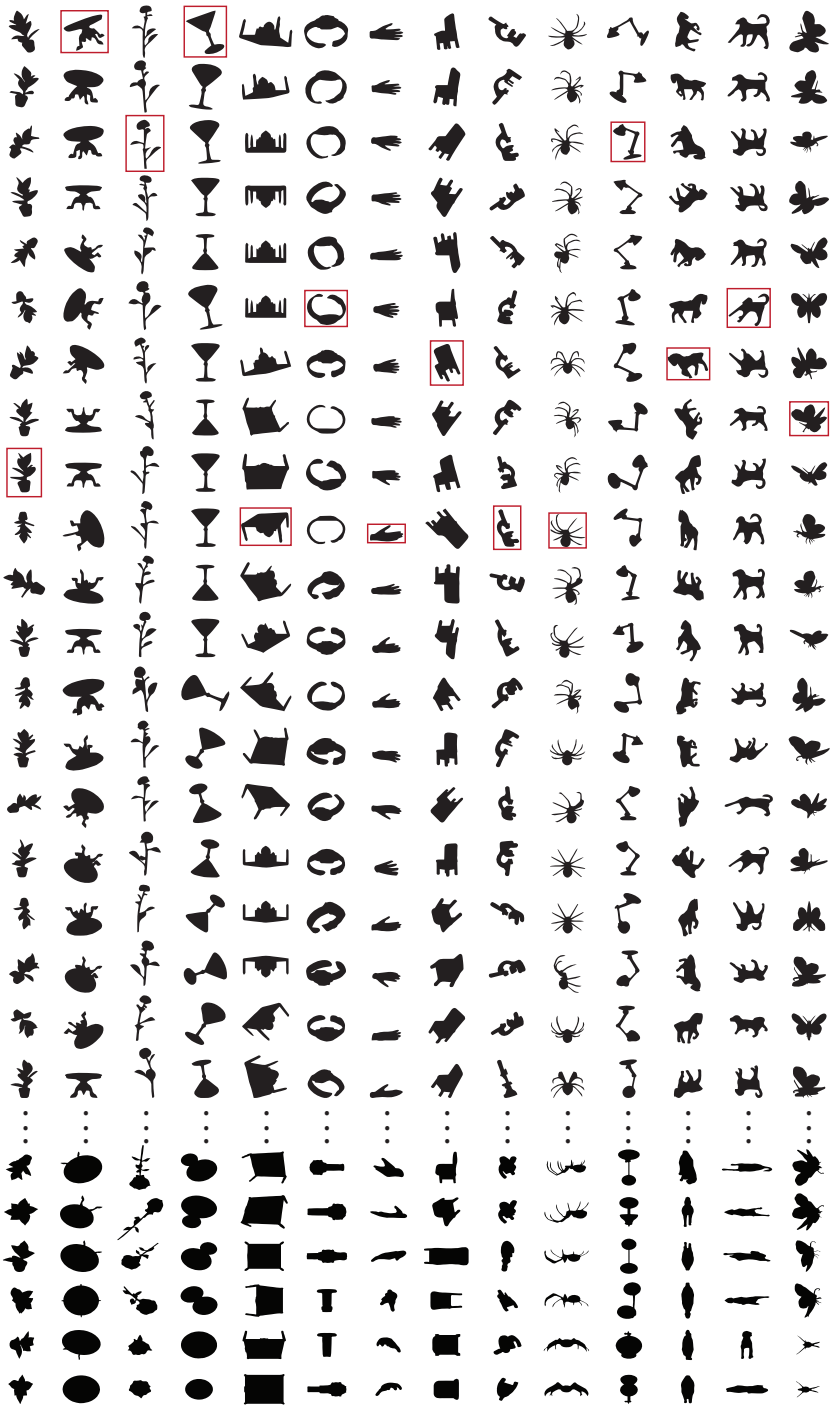


Figure 12: The silhouettes sorted in the decreasing order of skeletal lengths. The first 20 rows show those with the highest 20 SL values. The last 6 rows show the silhouettes with minimal SL where the last row represents the fiftieth view which is along the principle axis for most objects. The silhouette with minimal symmetry selected from the top 10 skeletal lengths is marked by the red box.

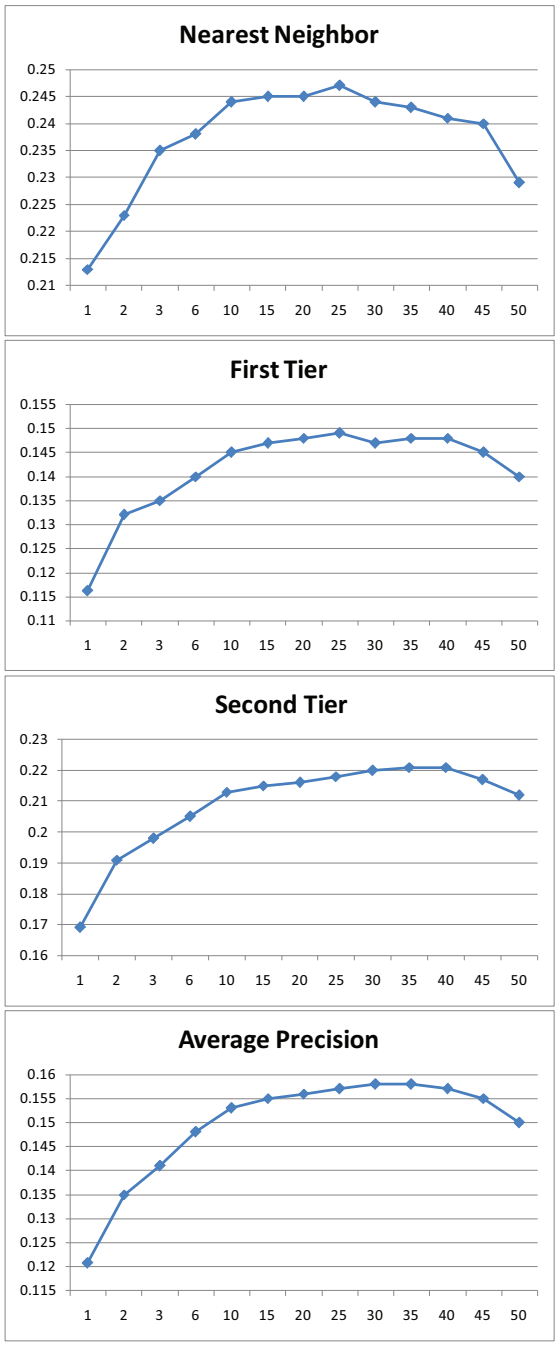


Figure 13: Performance metrics evolvement from 2 view representation up to 50 views.

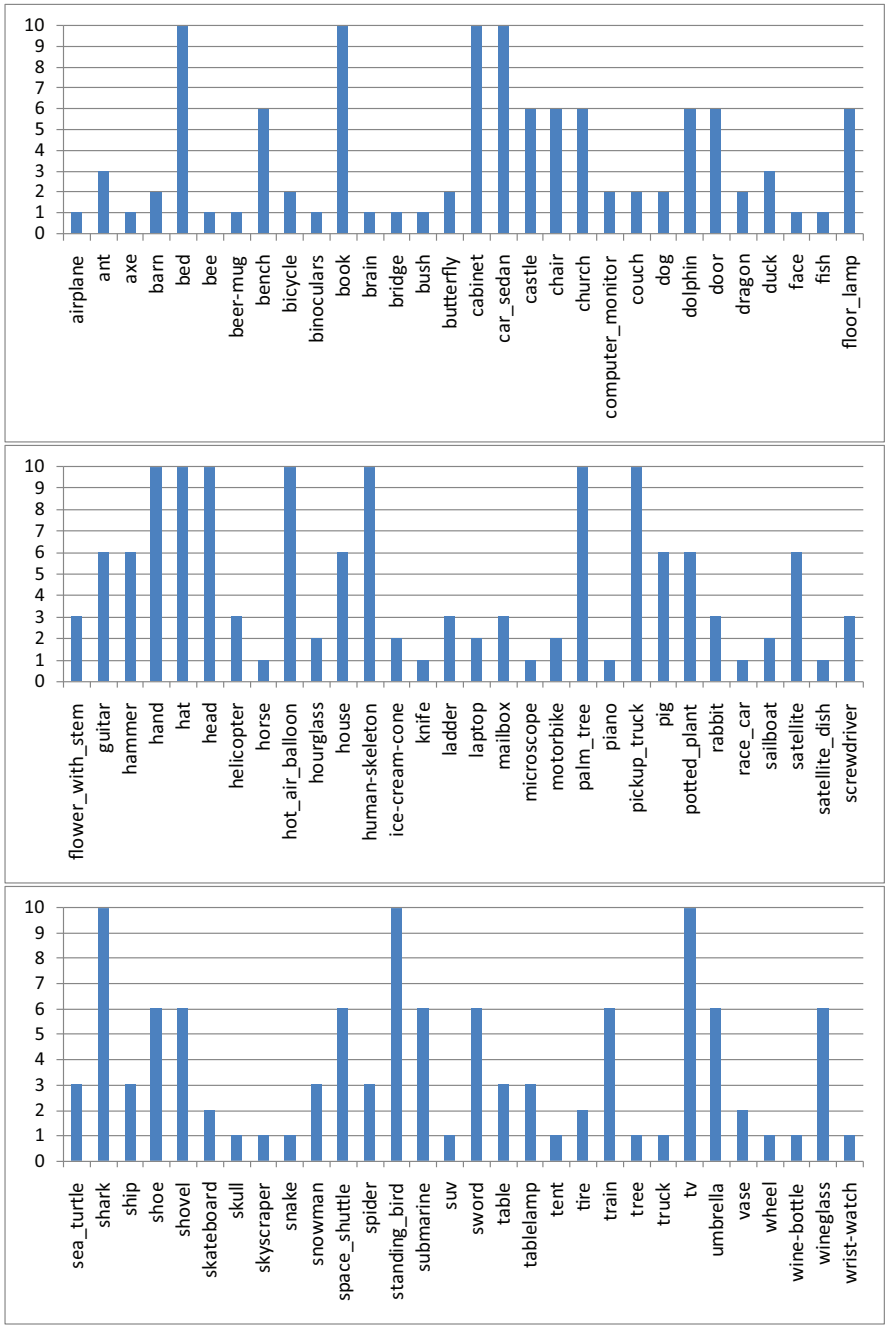


Figure 14: The number of views associated with the highest NN values for each class.

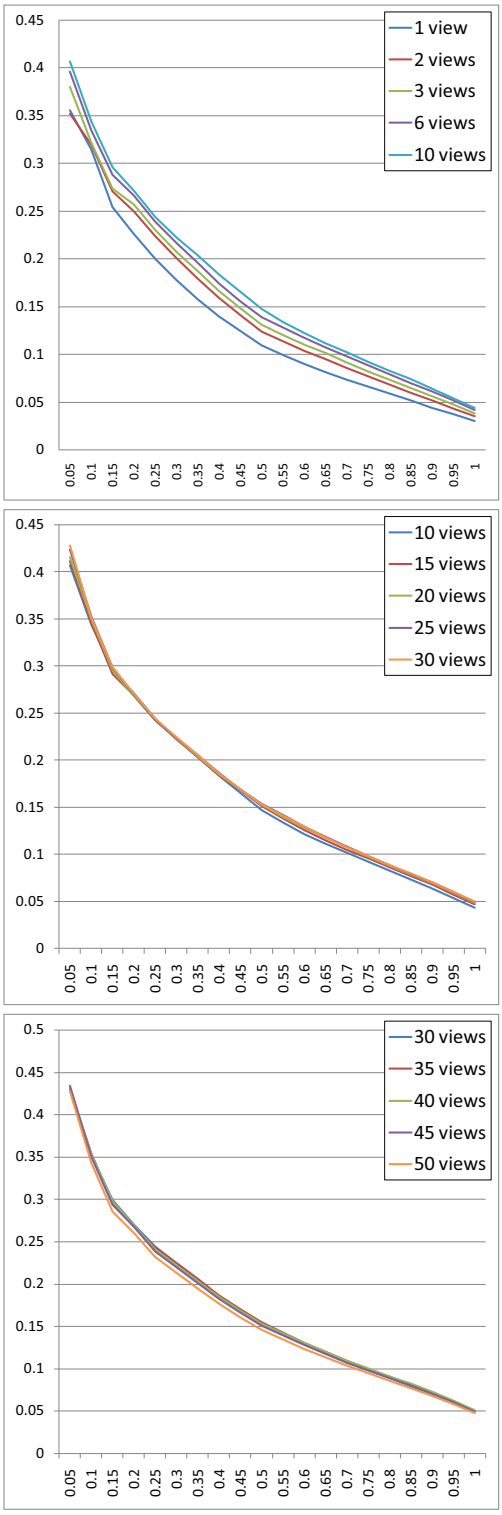


Figure 15: Precision Recall plot on target dataset where 3D models are represented by different numbers of views

6.1. Side Views

As concluded by the two learning methods [23, 40] mentioned in the introduction, the maximal information hypothesis had many counter-examples. Zhao el al. [40] refer it to a bias for front and side views when human draws an object. This contradicts the asymmetry hypothesis by Blanz et al. [4] to a certain extent since front views of symmetric objects are usually very symmetric. However, the side view preference observation in [40] can be explained by the fact that a subject with no artistic skills would describe a car, a horse, or a microscope from the side view violating the maximal information hypothesis in some cases such as the horse. In this section we look for candidate side views, show examples and test their effectiveness.

The side view criteria does not overthrow the maximal information criteria. The representative views set of a 3D object will always contain the maximal SL view. In addition to that view, we follow three different strategies to capture a side view. The first one which was originally proposed in [38] selects the off–axis view as the most asymmetric one from the top 10 (SL) silhouettes. The other two strategies exploit shape symmetry and the 3D object's principal axis.

While devising an alignment method for 3D models, Chaouch et al. [6] experimentally validate a coupling between the principal axes and the approximate reflection plane symmetry. We apply this idea by collecting 10 silhouettes with the highest symmetric values. Then, out of this collection, we select the first side view V^1_{sym} as the silhouette with maximal SL. The set of views orthogonal to V^1_{sym} is traversed and the other side view V^2_{sym} is identified by its maximal SL again (see Fig. 16 second and third column).

The third strategy relies on the assumption that the fiftieth silhouette carrying the least amount of information about the projected object is the closest to be aligned along its principal axis. We call this view the top view for simplicity of expression. The first side view V^1_{axis} is the one which has maximal symmetry amongst those orthogonal to the top view. Then, the second side view V^2_{axis} is selected such that it is orthogonal to both V^1_{axis} and the top view (see Fig. 16 fourth and fifth column).

We extracted the same performance metrics for these criteria and found that they are slightly outperformed by the first 2 and 3 maximal SL views respectively (see precision recall plots in Fig. 17). However, it is worth noting that the following classes obtained higher NN and average precision values for the side view representation: ant, bush, castle, church, dragon, face, helicopter, hot-air-balloon, microscope, pickup-truck, potted-plant, sea-turtle, shoe, suv, train, wheel, and wrist-watch.

6.2. Comparison with Other Methods

Table 1 shows that our approach outperforms the methods tested on the testing dataset of the SHREC'13 Sketch Track Benchmark except for Furuya et al. [12] who employed machine learning by cross–domain manifold ranking (CDMR). However, on the large scale dataset of SHREC'14 Sketch Track Benchmark our method reported better results than that of Furuya et al. [12]. In addition, the precision recall plots shown in Figs. 18 and 19 show that our method performs best amongst its peers. The average response times per query on the testing dataset of the SHREC'13 of our method are 18.22, 36.21, and 54.01 seconds taking 1, 2, and 3 views respectively (see Table 2). The average response time on the complete dataset of SHREC'14 is 142.28 seconds using the 3 views representation.

Compared to other methods that participated in these tracks, Saavedra et al [30] use the least number of sample views for a 3D model. They use the 6 orthogonal views (top, bottom, left, 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,

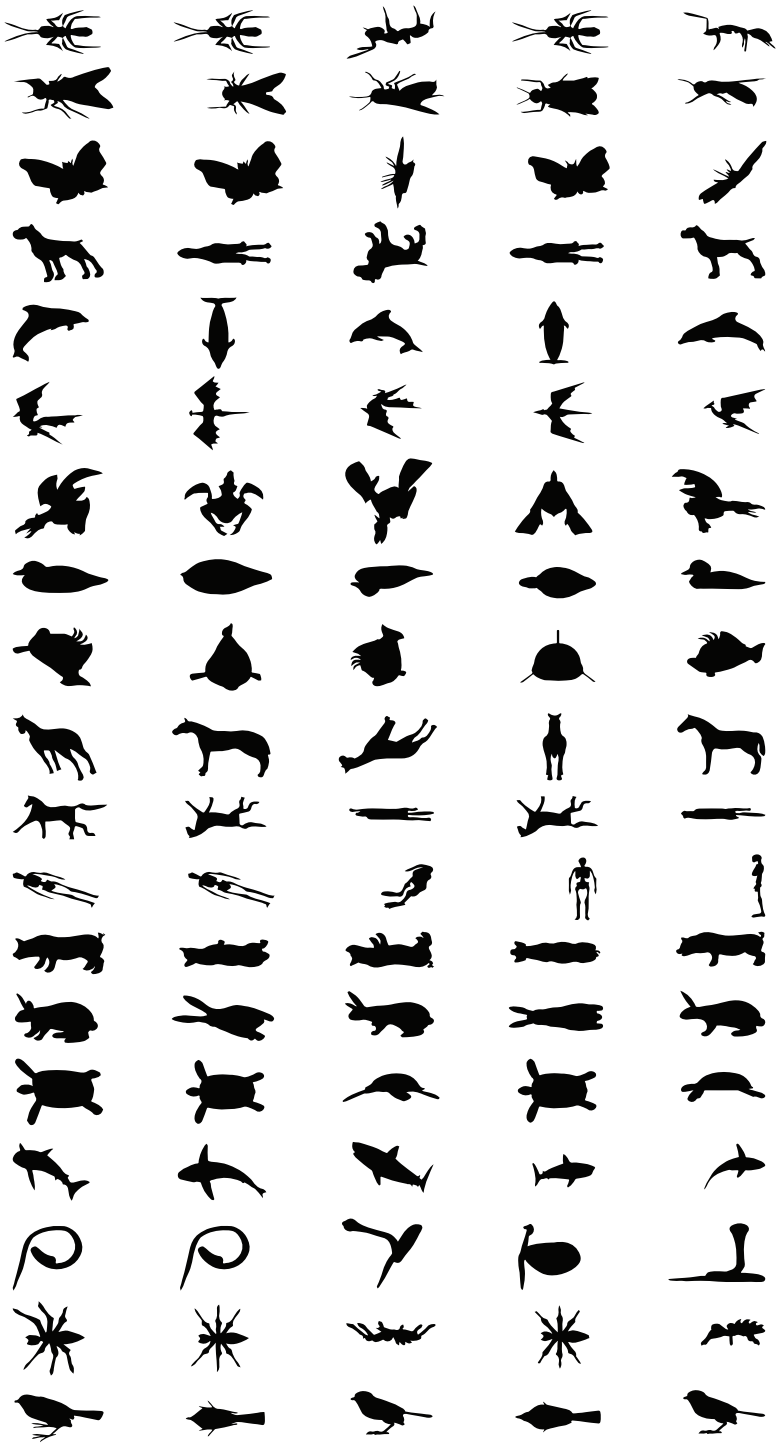


Figure 16: The side views extracted according to maximal symmetry criteria (columns two and three) and according to the principal axis criteria (columns four and five). The silhouettes with maximal SL are shown on the first column.

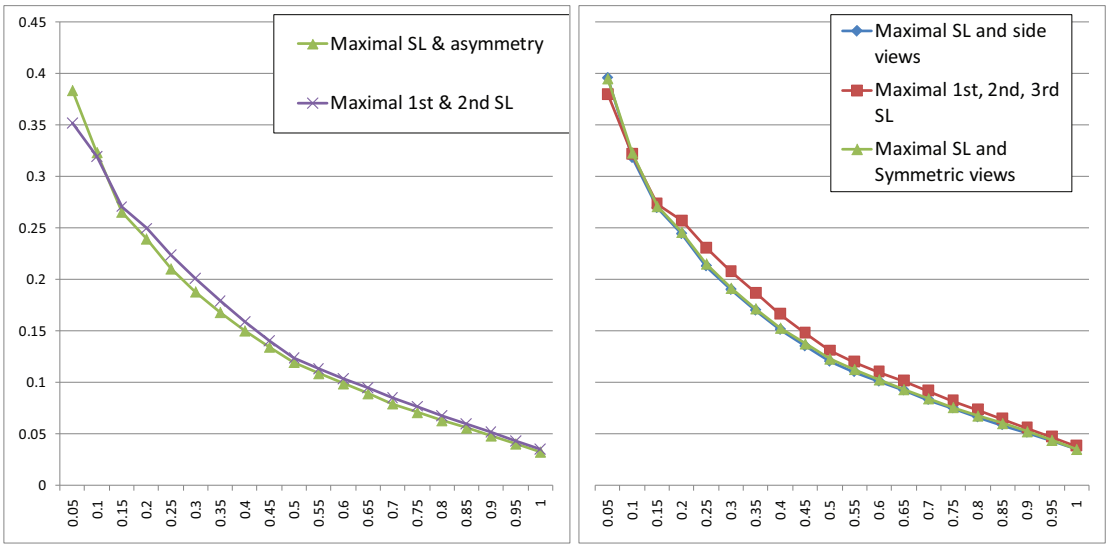


Figure 17: The precision recall plots of different runs using the same number of views. The plots show a superiority of views with maximal SL to those selected according to other criteria.

and consequently depicted, by humans. Despite increasing the number of views to 26 and 42 respectively, Aono et al. [2] and Zou et al. still score lowest on the precision recall plots. On the other hand, Li et al. [19] (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) [22], 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. Their participation in SHREC'14 included two runs with 9.5 ($\alpha = 0.5$) and 18.5 ($\alpha = 1$) views for each model (see Table 3). Furuya et al. [12] use the highest number of views proposed in this field (162 views) followed by Tatsuma (102 views) and still need machine learning to improve their retrieval results increasing the retrieval time in an enormous leap (0.49 seconds for BF-fGALIF to 615.95 seconds for CDMR-BF-fGALIF).

Reporting better performance over these methods while using a small number of sample views, we verify the merit of the "informative" criteria in viewpoint selection and 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 [16] on the testing dataset of the Sketch Track Benchmark and ranked second on the complete dataset of SHREC'14 [17]. 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 there are *incorrect* views for 3D models that cause misinterpretation and mismatching and thus *must* be eliminated from its set of sample views. We examined

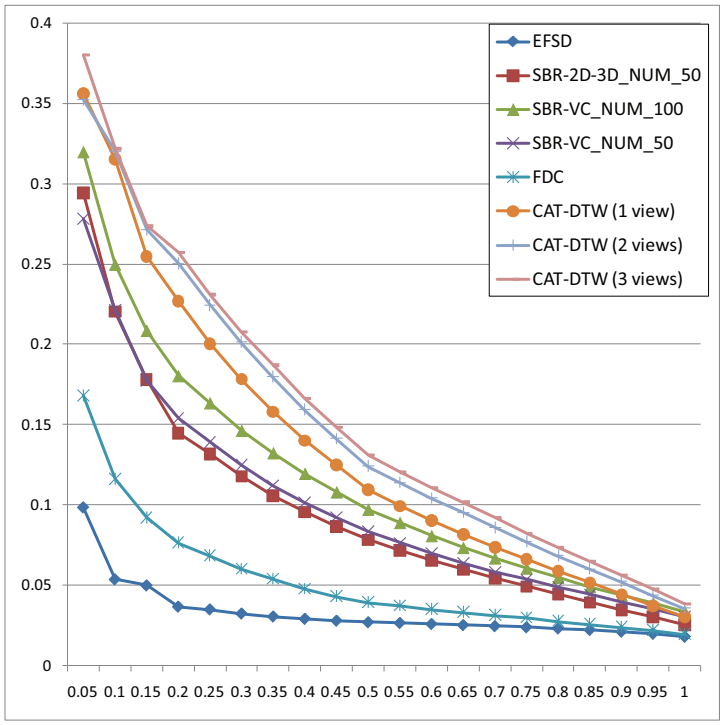


Figure 18: Precision–Recall diagram performance comparisons on the testing datasets of the SHREC'13 Sketch Track Benchmark.

PARTICIPANT	METHOD	NN	FT	ST	E	DCG	AP
Aono [2]	EFSD	0.023	0.019	0.036	0.019	0.24	0.031
Li[19]	SBR-2D-3D-NUM-50	0.132	0.077	0.124	0.074	0.327	0.0947
Li[22]	SBR-VC-NUM-100	0.164	0.097	0.149	0.085	0.348	0.1138
	SBR-VC-NUM-50	0.132	0.082	0.131	0.075	0.331	0.0984
Saavedra [30]	FDC	0.053	0.038	0.068	0.041	0.279	0.051
Furuya [12]	BF-fGALIF	0.176	0.101	0.156	0.091	0.354	0.119
	BF-fDSIFT	0.145	0.099	0.154	0.093	0.351	0.115
	CDMR-BF-fDSIFT	0.217	0.156	0.231	0.135	0.411	0.193
	UMR-BF-fDSIFT	0.154	0.113	0.178	0.104	0.366	0.133
	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 (1 view)	0.213	0.1162	0.1692	0.098	0.3713	0.1208
	CAT-DTW (2 views)	0.223	0.132	0.191	0.106	0.387	0.135
	CAT-DTW (3 views)	0.235	0.135	0.198	0.109	0.392	0.141

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

the asymmetry and the side view criteria through topological inversion and identification of the principal axis of an object and its orthogonal viewpoints respectively.

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 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, 36, 31, 15]. In addition to these methods, some pruning strategies can be applied to avoid the detailed correspondence computations for each configuration.

The side view criteria for view selection presents promising possibilities. Despite the slight superiority of the maximal SL views on the whole database, some classes report better results with the side views representation. Investigating these cases is work in progress. In addition, the set of candidate side views collected based on the symmetry and principal axis criteria is liable for further expansion and subsequent pruning.

8. Acknowledgements

We thank the Computer Science Department of the American University of Beirut for offering lab space and machines to perform the extensive tests presented in this paper. Particular thanks are due for Mr. Mustapha (Mike) Hamam, the systems analyst of the department.

Participant (with computer configuration)	Method	Language	t	R
Furuya (CPU: Intel Core i7 3930 K @ 3.20 Hz; GPU: NVIDIA GeForce GTX 670; Memory: 64 GB; OS: Ubuntu 12.04)	BF-fDSIFT	C++, CUDA	1.26	2
	BF-fGALIF	C++	0.49	2
	CDMR-BF-fDSIFT	C++	606.96	9
	CDMR-BF-fGALIF	C++	615.95	9
	UMR-BF-fDSIFT	Matlab	54853.77	10
	UMR-BF-fGALIF	Matlab	27219.49	10
Li (CPU: Intel Core 2 Duo E7500 @ 2.93 GHz; Memory: 16 GB; OS: Windows 7 64-bit)	SBR-VC-NUM-100	C/C++	208.85	8
	SBR-VC-NUM-50	C/C++	68.92	7
	SBR-2D-3D-NUM-50	C/C++	43.93	5
Saavedra (CPU: Intel Core i7-3770 CPU @ 3.40 GHz; Memory: 8 GB; OS: Ubuntu 11.10)	HOG-SIL	C++	0.09	1
	HELO-SIL	C++	0.09	1
	FDC	C++	0.09	1
Our Method (CPU: Intel(R) Core(R) CPU ES-2650 0 @ 2.00 GHz 2.00 GHz (2 processors); Memory: 32.0 GB; OS: Windows 7 64-bit	CAT-DTW (1 view)	C++	18.22	3
	CAT-DTW (2 views)	C++	36.21	4
	CAT-DTW (3 views)	C++	54.01	6

Table 2: Timing information comparison on the testing dataset of SHREC'13 Sketch Track Benchmark: t is the average response time (in seconds) per query. "R" denotes the ranking group number.

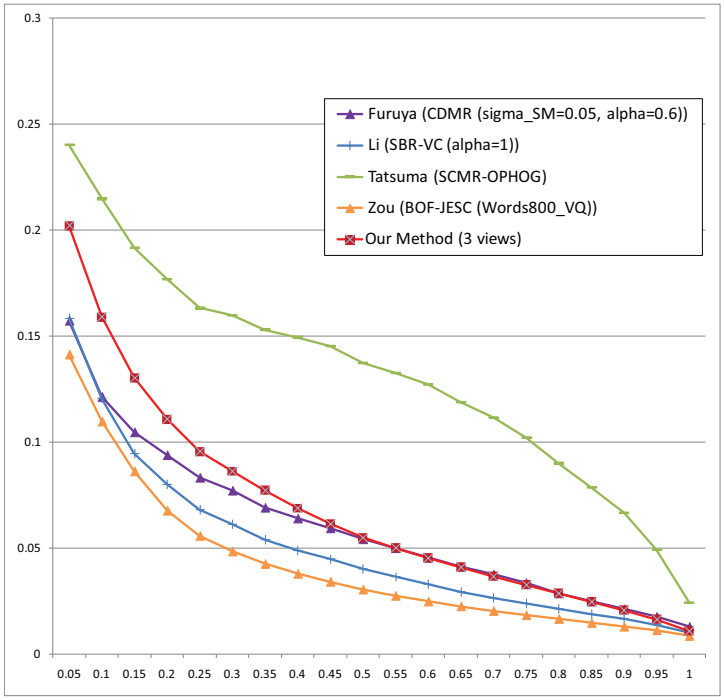


Figure 19: Precision–Recall diagram performance comparisons of the best runs on the complete datasets of the SHREC'14 Sketch Track Benchmark.

PARTICIPANT	METHOD	NN	FT	ST	E	DCG	MAP
Furuya	BF-fGALIF	0.114	0.05	0.079	0.036	0.321	0.045
	CDMR (sigma_SM=0.05, alpha=0.3)	0.109	0.057	0.09	0.041	0.329	0.054
	CDMR (sigma_SM=0.05, alpha=0.6)	0.084	0.058	0.094	0.04	0.325	0.06
	CDMR (sigma_SM=0.1, alpha=0.3)	0.102	0.055	0.087	0.039	0.324	0.053
	CDMR (sigma_SM=0.1, alpha=0.6)	0.068	0.046	0.074	0.031	0.308	0.048
Li	SBR-VC (alpha=0.5)	0.09	0.047	0.077	0.035	0.316	0.046
	SBR-VC (alpha=1)	0.096	0.05	0.081	0.038	0.319	0.05
Tatsuma	OPHOG	0.159	0.066	0.098	0.051	0.341	0.061
	SCMR-OPHOG	0.158	0.117	0.171	0.078	0.376	0.132
Zou	BOF-JESC (FV_PCA32_Words128)	0.095	0.039	0.061	0.027	0.303	0.037
	BOF-JESC (Words1000_VQ)	0.094	0.039	0.063	0.028	0.306	0.039
	BOF-JESC (Words800_VQ)	0.099	0.043	0.068	0.031	0.311	0.042
Our Method	CAT-DTW (1 view)	0.12	0.06	0.091	0.043	0.329	0.051
	CAT-DTW (2 views)	0.13	0.065	0.099	0.048	0.335	0.057
	CAT-DTW (3 views)	0.137	0.068	0.102	0.05	0.338	0.06

Table 3: Performance metrics for the performance comparison on the complete dataset of the SHREC'14 Sketch Track Benchmark.

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] Marco Bertamini and Johan Wagemans. Processing convexity and concavity along a 2-d contour: figureground, structural shape, and attention. *Psychonomic Bulletin & Review*, 20(2):191–207, 2013.

[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] Mohamed Chaouch and Anne Verroust-Blondet. Alignment of 3d models. *Graphical Models*, 71(2):63–76, 2009.

[7] Elias H. Cohen and Manish Singh. Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research*, 47(22):2825 – 2840, 2007.

[8] J. De Winter and J. Wagemans. The awakening of attneave's sleeping cat: Identification of everyday objects on the basis of straight-line versions of outlines. *Perception*, 37:245–270, 2008.

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

[10] Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? *ACM Trans. Graph.*, 31(4):44:1–44:10, July 2012.

[11] 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.

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

[13] Donald D Hoffman and Manish Singh. Salience of visual parts. *Cognition*, 63(1):29 – 78, 1997.

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

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] Peter Neri. Wholes and subparts in visual processing of human agency. *Proceedings of the Royal Society B: Biological Sciences*, 276(1658):861–869, 2009.

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

[33] 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.

[34] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser. The Princeton shape benchmark. In *Shape Modeling International*, June 2004.

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

[36] 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.

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

[38] Zahraa Yasseen, Anne Verroust-Blondet, and Ahmad Nasri. Sketch-based 3D Object Retrieval Using Two Views and a Visual Part Alignment. In I. Pratikakis, M. Spagnuolo, T. Theoharis, L. Van Gool, and R. Veltkamp, editors, *3DOR 2015 - Eurographics Workshop on 3D Object Retrieval*, page 8, Zurich, Switzerland, May 2015.

[39] 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.

[40] Long Zhao, Shuang Liang, Jinyuan Jia, and Yichen Wei. Learning best views of 3d shapes from sketch contour. *The Visual Computer*, pages 1–10, 2015.

[41] 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.

[42] Changqing Zou, Changhu Wang, Yafei Wen, Lei Zhang, and Jianzhuang Liu. Viewpoint-aware representation for sketch-based 3d model retrieval. *Signal Processing Letters, IEEE*, 21(8):966–970, 2014.