Qualitative part-based models in content-based image retrieval

Guillaume-Alexandre Bilodeau[†*] and Robert Bergevin[‡]

[†]*Department of Computer Engineering, École Polytechnique de Montréal*
P.O. Box 6079, Station Centre-ville, Montréal (Québec), Canada, H3C 3A7

[‡]*Computer Vision and Systems Laboratory, Pavillon Adrien-Pouliot, Université Laval, Ste-Foy (Québec), Canada, G1K 7P4*

Machine Vision and Applications

[*] Corresponding author. Tel: +1 514 340 4711 x5064 fax: +1 514 340 3240
E-mail address: guillaume-alexandre.bilodeau@polymtl.ca

Abstract

A qualitative, volumetric part-based model is proposed to improve the categorical invariance and viewpoint invariance in content-based image retrieval, and a novel two-step part-categorization method is presented to build it. The method consists first in transforming parts extracted from a segmented contour primitive map and then categorizing the transformed parts using interpretation rules. The first step allows noisy extracted parts to be transformed to the domain of a simple classifier. The second step computes features of the transformed parts for categorization. Content-based image retrieval experiments using real images of complex multi-part objects confirm that a model built from the categorized parts improves both the categorical invariance and the viewpoint invariance. It does so by directly addressing the fundamental limits of low-level models.

Keywords

Part-based model; image retrieval; qualitative volumetric primitives; shape categorization.

1. Introduction

Consider the two desk lamp images in Fig. 4. Their pixels have different intensities and their backgrounds differ. Now, suppose that a content-based image retrieval system needs to retrieve the second image, given the first as a query. These images seem simple, but, clearly, any approach based on colors, textures, 2D shapes or any statistic on the pixels will fail, since this information is low-level and not characteristic of the desk lamp in the images. The fact that humans can match these desk lamps is due to their use of a higher-level interpretation process.

One way to match these two desk lamps is through qualitative, volumetric part-based models. This approach was proposed in the 1980s, and studied mostly in the 1980s and at the beginning of the 1990s (Bergevin and Levine, 1993, Biederman, 1985, Dickinson et al., 1992). These models were said to allow viewpoint invariance. They did, theoretically, but difficulties with obtaining the parts from images made them unpopular. Still, their theoretical benefits are appealing for many computer vision tasks, including generic object recognition, content-based image retrieval, and even video surveillance.

Dickinson et al. (1998) successfully used a qualitative part-based model in content-based image retrieval by searching individual parts of a single target object (a lamp) in a complex scene. Here, we push this idea, in two different directions: categorical invariance and viewpoint invariance. First, no target object is assumed. A query image of a main unknown object (the system knows only that the object is potentially multi-part) is processed in order to extract a complete qualitative volumetric part-based model of that object. This model is then compared to database models similarly obtained from other images in order to retrieve and sort images with similar contents, irrespective of color, texture, background, viewpoint, and within-category object shape variations. For instance, a white desk lamp should be more similar to a blue table lamp than to a white airplane. Also, identical lamps seen from quite different viewpoints should be more similar than different lamps seen from the same viewpoint. Even though recent appearance-based approaches, such as the one proposed by Dorko and Schmid (2003), offer significant improvement towards categorical invariance, they still fall short of providing the needed degree of invariance with respect to viewpoint, color, texture, and background. Furthermore, the notion of object is more precise in our case: a multi-part object the parts of which are volumes of simple shape. In appearance-based approaches, the notion of object is fuzzier. For instance, in Sivic and Zisserman (2003), an object is the contents of any selected image window. In Schaffalitzky and Zisserman (2002), an object is a complete outdoor scene. Hence, these methods are not designed for a query based on object identity.

Part-based models using bottom-up reconstruction of generalized cylinders from image features have been investigated by a number of researchers. Methods

proposed by Gross and Boult (1996), Liu et al. (1993), Ponce et al. (1989), Sato and Binford, (1993), and Zerroug and Nevatia (1999) detect symmetry axes and cross sections, and use rules based on invariant and quasi-invariant properties to recover parametric models of specific types of generalized cylinders. A generic method covering a larger set of possible object shapes would be needed, but this is a difficult task that has not yet been accomplished.

Liu et al. (2002) and Pilu and Fisher (1996) proposed instead a top-down approach, where a limited number of parameterized volumetric primitives are projected on the parts and deformed in such a way that an error measure is minimized. As explained below, such an analytical fitting process is not possible with the qualitative part-based models needed here.

Generic bottom-up extraction and interpretation of object parts from contour primitives has also been attempted (Bergevin and Levine, 1993, Dickinson et al., 1992, Jacot-Descombes and Pun, 1997). Hierarchical grouping processes produce contour primitives, faces, volumetric parts, and objects. This approach was initially tested on simple or synthetic images due to the difficulty of extracting the proper contour primitives from complex images. Bilodeau and Bergevin (2002) presented an algorithm simplifying the grouping processes when an object outline is available. Original algorithms were proposed to extract a multi-part object outline from a constant-curvature contour primitive (CCP) map, and then to extract object parts on that basis. Viewpoint invariance was shown to be feasible at the part level. Now, in order to push the idea further, the obtained parts are categorized and combined to form qualitative part-based models of objects. These models are matched in the context of content-based image retrieval.

In this paper, a new method is proposed, as part of the model-building process, which is aimed at interpreting segmented parts using simple qualitative volumetric primitives. The method consists of part transformation, followed by a rule-based categorization of the transformed parts. The first step allows the transformation of any possible part in order to be in the domain of a simple classifier. The second step computes features of the transformed parts for categorization. Combining categorized parts results in a qualitative part-based model that is useful for image retrieval. An experimental image retrieval system, called PLASTIQUE, was developed on the basis of this new method. The results obtained show that a qualitative volumetric part-based model adds partial categorical and viewpoint invariance to image retrieval. For instance, PLASTIQUE retrieves the second image in Fig. 4, given the first as a query.

In section 2, major features of the part segmentation algorithm are reviewed. Section 3 explains the proposed method for part categorization. Section 4 describes briefly how categorized parts are used and compared. In section 5, the qualitative volumetric part-based model obtained is validated in image retrieval experiments. Finally, in section 6, we conclude that image retrieval can benefit from a robust categorization method combined with part segmentation and a fuzzy matching method.

2. Parts from a CCP map and an object outline

In order to be transformed and categorized, parts must first be extracted from the image. A constant-curvature contour primitive (CCP) map (see Fig. 4) is obtained by processing a grey-level image with a generic contour extraction and a segmentation algorithm (Mokhtari and Bergevin, 2001). This algorithm is based on a standard Canny edge detector, a contour extraction method, and an original

contour segmentation and approximation method. This last method enforces a number of generic shape criteria. Each *CCP* is meant to be limited to a single object or volumetric part, and *CCP*s can be grouped efficiently into parts. CCP grouping processes which are robust to partial misses should be able to produce an object outline and parts. Previous contour-based perceptual grouping methods have been proposed to build object faces (Bergevin and Levine, 1993, Dickinson et al., 1992, Jacot-Descombes and Pun, 1997), local contour structures (Forsyth et al., 1998, Selinger and Nelson, 1999), or convex groups (Jacobs, 1996). However, most of these methods, notably those proposed by Bergevin and Levine (1993) and Dickinson et al. (1992), assumed nearly perfect CCP maps and relied on internal edges, which are difficult to identify in real images.

The extraction of generic part outlines that are constrained by the object outline was proposed by Bilodeau and Bergevin (2002). Constraints are associated with the way parts arise from protrusions and indentations in a multi-part object. They are related, but simpler than the minimum-cut rule of Singh et al. (1999). Parts are extracted on the basis of CCP pairs optimizing a perceptual grouping criterion. The outline of the object is used as a structural clue.

Obtaining the object outline or silhouette is known as the figure-ground segmentation problem. This is still a major challenge, especially for static scenes. Numerous methods have been published, many of which are limited to finding image contours with no explicit reference to meaningful objects. Work is still under way to address this fundamental machine vision problem (Martin et al, 2004, Elder and Zucker, 1996, Hérault and Horaud, 1993, Randrianarisoa et al., 2005, Bernier and Bergevin, 2006). Meanwhile, most shape analysis and interpretation methods assume a nearly perfect figure-ground segmentation

(Mokhtarian, 1996). In this work, such an unrealistic assumption is not made. The experiments in this paper use a generic figure-ground segmentation algorithm based on a backtracking graph search of closed CCP paths from multiple starting primitives (Bilodeau and Bergevin, 2000). The outlines obtained are typically quite good for test images of close-up views of single complete objects with highlights, shadows, internal texture and markings, and moderate background clutter. In cases where the background clutter is more dominant, e.g. the airplane image in Fig. 4, some spurious structures may be appended to the outline. In other cases, an incomplete or noisy outline may be obtained. However, in all cases, the perturbations are local.

In contrast, part extraction methods based on region analysis and symmetry transforms (Siddiqi et al., 1999) are likely to be more globally affected by significant missing or spurious silhouette pieces. Such significant errors occur when existing figure-ground segmentation methods are applied to realistically complex images, as in the present work. However, a necessary condition for success in the interpretation stage of a part-based method is that only local errors be allowed. In order to meet this condition, our proposed method includes further processing steps designed to provide local errors with the needed robustness. As a last resort, a limited set of well-identified parts may very well be sufficient for successful matching (Biederman, 1985).

Figure 1

In Fig. 1, a flow diagram of the processing steps leading to part segmentation is presented (Bilodeau and Bergevin, 2002). The object outline and each part are made up of an ordered and closed list of contour primitives selected from a CCP map. Exterior CCP removal consists in removing CCPs that are not within the

region enclosed by the outline. Grouping attempts are made by forming pairs of CCPs based on parallelism, proximity, similarity in length and type, and regional overlap. Two types of grouping attempts are made. The first uses the outline as a structural grouping clue. The second does not. Hence, the first type of grouping depends on, and is restricted by, the extracted outline, whereas the second type makes pairs of CCPs, regardless of the outline (for internal parts and outline errors). Grouped CCPs are considered as two main sides of a part. Boundary completion consists in completing the outline of a part by adding CCPs from the map to make a close contour. Finally, group removal consists in removing from the CCP map those CCPs corresponding to the detected part.

Examples of part segmentations are presented in Fig. 4. The parts obtained cover most of the region delimited by the extracted outline, and overlap between parts is limited. The obtained parts usually represent a meaningful, though noisy, structural decomposition of the object. Coherent parts are obtained over a large range of viewpoints (Bilodeau and Bergevin, 2002). This will be discussed further in section 5.1.

3. Part-based categorization

Two processing steps are required in order to obtain qualitative volumetric parts from segmented projected object parts: part transformation and part categorization. Parts are categorized according to a set of eighteen qualitative volumetric primitives (see Fig. 2). Volumetric primitives must be limited to simple shapes, since parts are extracted from pairs of CCPs approximating projected sides.

Figure 2

Geons (Biederman, 1985) are a natural choice for simple qualitative volumetric primitives, but this set was reduced by taking away primitives arising from

different cross-section symmetries, as this attribute is hard to compute in real images of multi-part objects. We recall here that qualitative volumetric primitives have no unique or parameterized specific shape. Instead, they are defined by a limited set of qualitative shape attributes. Therefore, no analytical representation of the volumetric primitives or of their projections could be fitted to the extracted parts. Instead, the defining attributes need to be related to the part CCPs, as will be the case in the second processing step. It is essential to keep in mind the scope of the proposed method and the generic nature of the interpretation process. Images of categorical objects (e.g. lamps or airplanes) must match, despite variations in scale, viewpoint, color, texture, background, and, most importantly, large within-category shape variations. Pixel-level silhouette matching, including the variants with scaling, rigid transformations, or even local linear deformations, are not designed for this task.

For efficiency reasons, the classifier must have a small number of rules, despite the fact that parts extracted from images have many different shapes due to within-category shape variations, viewpoint, and processing noise. This requires that parts be transformed into a common format prior to categorizing them. The following principle was formulated based on experiments: outlines of projected volumetric primitives may be simplified using a path made up of three or four constant-curvature contour primitives (CCPs) and still allow recovery of the primitive label, although with some ambiguity. Hence, in accordance with this principle, extracted parts are transformed into such simplified parts before being categorized into one or many qualitative volumetric primitive hypotheses.

3.1 Part transformation

Fig. 3 shows a flow diagram of part transformation and its position in the whole modeling process. The CCPs on the part outline are sampled to obtain dominant

points, a dominant point arising at a CCP endpoint with an abrupt change in orientation (greater than 30°, as determined experimentally) or curvature, and at the midpoints of circular arcs (see Fig. 3).

Figure 3

Sampling usually produces about ten to fifteen dominant points. To obtain a simplified part, four dominant points must be selected. These are the corners of a polygon with an enclosed area as similar as possible to the original part area, and as rectangular as possible. Given the initial set of dominant points, possible polygons are evaluated and the optimal one is selected according to

$$SPoly = \min_{poly_i \in polys} \left(\sum_{SP_i \in poly_i} (\angle SP_i - 90°) + \left| \frac{Area(OSP) - Area(poly_i)}{Area(OSP)} \right| \right), \qquad (1)$$

where SP_i is a sampled point and OSP is the ordered set (based on a clockwise sampling) of all the SP_i. $\angle SP_i$ is the angle formed by the vector made up of SP_i and the previous point in OSP, and the vector made up of SP_i and the following point in OSP. $poly_i = \{SP_a, SP_b, SP_c, SP_d\}$ is an ordered set of four points from OSP and $Area()$ is an operator evaluating the enclosed area of an ordered set of points. $polys = \{poly_i\}$ is the set of all the possible $poly_i$. The optional reduction to three CCPs is carried out afterwards by eliminating nearby points.

The polygon obtained is only a crude approximation for parts with circular arcs. It is modified to have sides that are circular arcs instead of straight-line segments if this produces a better fit. This is achieved through a pairwise match between the original contour sections and the polygon sections. Finally, nearby simplified parts may be merged according to similar criteria. Fig. 4 presents maps of simplified parts.

Figure 4

3.2 Simplified part categorization

A rule-based classifier was developed to categorize simplified parts. The classifier rules are based on six attributes obtained from the CCPs of the part and their geometric relationships. They are composed of a premise involving different attributes of the simplified part and a conclusion listing the possible qualitative volumetric primitives from which the projected part may arise. A ranking value is associated with each possible primitive to sort them according to their plausibility. The rules and the ranking values were set manually. Having many hypotheses contributes to viewpoint invariance by accounting for different views of the volumetric primitive category. Later on, when matching categorized parts, all hypotheses can be compared to find the best match by assuming that an object can be observed from different viewpoints with changes in the projections of its parts. The choice of values is not critical, as long as the ordering corresponds to the likelihood that a given projection arises from a given volumetric primitive. However, coherent rank values must be used among the various rules. The exact impact of the choice of rank values has not been thoroughly investigated, but initial tests have shown that modifying the values slightly has no significant impact. This results from using a fuzzy matching algorithm to assess model similarity (see section 4).

There are forty-nine rules, one for each possible simplified part type (see Table 1).

Table 1

The six attributes of the simplified parts are the number of CCPs (nCCPs), the number of circular-arc CCPs (nCAs), the convexity of CCPs (Conv), and three attributes of an axis-based description of the simplified part: straight-line segment parallelism (pSLS), circular arc compatibility (cCA), and the sweeping rule (Sweep). The expressions for the last four appear in the Appendix.

4. Use and comparison of categorized parts

In content-based image retrieval and in object recognition, categorized parts are combined in a graph to account for object structure. Each part (containing multiple volumetric primitive hypotheses) is a graph node, and proximity relationships between parts form the graph edges. Our approach to comparing these graphs is to use structural indexing (Nishida, 2002, Stein and Medioni, 1992). Instead of attempting to compare graphs pairwise, their level of similarity is established by a voting mechanism and by comparing subgraph structures. We redefined structural indexing in order to apply it to graphs with fuzzy attributes. The subgraphs are composed of either one node, two nodes and an edge, or three nodes and two edges. Compound hypotheses are devised for subgraphs of more than one node. Prior to being compared, the subgraphs are sorted according to their entropy. A comparison is made by matching subgraphs with low entropy (less ambiguous) to subgraphs with high entropy (many hypotheses). The most plausible volumetric primitive hypotheses are first compared. After testing the best hypotheses for each subgraph, those remaining are compared using their second-best hypothesis, and so on.

A matching score is attributed each time subgraphs are matched. This is done by adding the minimum values of each non-zero hypothesis; formally, the matching is

$$MS = \frac{\sum_{i \approx j} \min(RV_Q(i), RV_D(j))}{\sum RV_Q(i)}, \qquad (2)$$

where $RV_Q(i)$ and $RV_D(j)$ are the ranking values for the ith or jth hypotheses of subgraph Q and D respectively. Symbol $i \approx j$ means that the sum is computed only

for matching hypotheses in the two subgraphs. The total score for two graphs is given by

$$TS = a\sum_{i=1}^{3}\left(\frac{SGEQ_i}{SGEDB_i}\right)+\sum_{i=1}^{3}\left(\frac{b_i\sum MS_i}{\min(SGEQ_i, SGEDB_i)}\right), \quad (3)$$

where $SGEQ_i$ and $SGEDB_i$ are the numbers of subgraphs with i nodes in the two compared graphs. MS_i is a matching score for a subgraph of i nodes, and a and b_i are weighting coefficients. The four weighting coefficients are adjusted dynamically: the smaller the term a coefficient multiplies, the larger its adjusted value. This allows the differences between the two graphs to be emphasized. The value for *TS* varies between 0 and 1, where 1 is a perfect match. More details can be found in Bilodeau and Bergevin (2005).

5. Validating experiments

In this section, categorical invariance and viewpoint invariance are tested, along with our two-step categorization method. Two experiments are performed: a clustering experiment and an image retrieval experiment. About 1300 parts extracted from 312 images were categorized and used in our experiments. A sample of images and their simplified parts is shown in Fig. 5.

5.1 Clustering experiment

This experiment was designed to assess the ability of our method to interpret and match sets of noisy parts. Those parts are extracted from images depicting instances of six object types observed from different viewpoints. Three of the object types (coffee cup, lamp, and airplane) include objects with significant within-category variations. Viewpoint invariance is also an important factor, since projected silhouettes may be quite different in distant viewpoints, even for the same object instance.

The clustering experiment task consists in making up groups of objects of the same category on the basis of their extracted parts. The performance of the proposed method is compared with human performance under similar conditions, both assessed with respect to ground truth categories. In this way, the effect of part segmentation errors, which are not linked to part transformation and categorization but still have an impact on clustering performance, is removed. The proposed part categorization method is also compared with a modified version that categorizes simplified parts only by their 2D shapes and deformations (rectangle, triangle, trapezoid, disk, etc.). Hence, a rectangle and a trapezoid will be related, but not a rectangle and a circle (both views of a cylinder). In this way, it will be possible to assess whether or not explicitly considering hypotheses of volumetric primitives and viewpoints is useful.

Graphs of volumetric primitives and fuzzy graph matching, as described in section 4, are used in this experiment. The automatic clustering is conducted in the following manner. A query image is selected at random from the database. All images with models that are sufficiently similar to the query image model are grouped together and removed from the database. Then, another query image is selected from the remaining images in the database, and so on, until the database is empty. We performed this procedure with similarity thresholds of 0.6 and 0.7 (perfect similarity being equal to 1). Threshold values larger than 70% result in too many small groups of images, while values smaller than 60% result in too few large clusters.

Six people, four of them familiar with computer vision (S1, S4, S5 and S6 in Table 2) were asked to perform a similar clustering task. Printouts of the simplified part maps of 195 images (see Fig 5 for several examples) were

presented to each person in turn, and each was asked to group together the printouts corresponding to each object type. That person was also told that the objects could be seen from different viewpoints and could present within-category shape variations. The number of groups (clusters) was not specified.

Figure 5

In order to compare the automatic and human performances, measures of purity and entropy were used, as defined below (Dhillon et al., 2001). Cluster C_i's purity can be defined as

$$P(C_i) = \frac{1}{n_i} \max_h (n_i^{\ h}) , \tag{4}$$

where $n_i = |C_i|$ and $n_i^{\ h}$ is the number of images that belong to the object category h, h=1, …,j. A cluster may contain samples from different object categories. Purity gives the ratio of the dominant object category size in the cluster to the cluster size itself. Since an object category can dominate more than one cluster, we computed the purity for an object category by adding (with a weighting based on the cluster size) the purity of all clusters where the category dominates. A high purity value means that the cluster is a pure subset of a dominant category.

We also used entropy, which is defined by

$$H(C_i) = -\frac{1}{\log\ j} \sum_{h=1}^{j} \frac{n_i^h}{n_i} \log\left(\frac{n_i^h}{n_i}\right). \tag{5}$$

Entropy considers the distribution of object categories in a cluster. Entropy has been normalized to take values between 0 and 1. An entropy of 0 means that the cluster contains only one object category, while an entropy near 1 implies that the cluster contains a uniform mixture of object categories. We also computed the number of clusters dominated by each object category and the number of printouts

not clustered (i.e. alone in a cluster). Table 2 gives the results for these four measures.

Table 2

First, let us compare our method with the modified version that considers only 2D shape deformations. Using only 2D shapes results in poorer clustering performance. More clusters per category means that fewer images in a category are considered to be alike. This is reflected in the number of images that were not clustered. Furthermore, purity and entropy values are lower, as more partial graph matches are performed by the matching algorithm. These results demonstrate that considering multiple hypotheses based on the viewpoint of volumetric primitives allows better matching performance.

Comparing human performance with our method, we note that there is a large inter-person variability for some objects due to noisy simplified part maps. If we first consider the number of clusters for each object category, human subjects made an average of three clusters per category. This means that each object category can be described in an average of three ways. This is an indicator of the stability of the extracted part sets for each object category. Although not perfectly stable, the objects are grouped into a small number of descriptions. On this aspect, humans outperformed our method, but the number of possible descriptions is still limited. Another aspect is the number of printouts that are not clustered. This indicates descriptions that did not have much in common with the others. In this case, we can conclude that these are segmentation failures. Overall, human subjects consider (indirectly) that about 9% of printouts have been subject to segmentation failures, which is equivalent to our method using a 60% threshold. This percentage is doubled if a 70% threshold is used. The latter results in stricter

discrimination between descriptions (more clusters per category) and better purity and entropy. There is a trade-off to be had between the number of clusters and their purity (a small number of large clusters often means a low level of purity). For image retrieval, the 70% threshold is a good choice for our proposed method.

As for purity, the clusters made by human subjects are better, although some printouts may still be too imperfect to enable a suitable identification by humans. For instance, printouts for some views of the coffee cup (Subject 4 failed to make any cluster dominated by the coffee cup category) and the stool led to a lower level of purity than did those of the other object categories. One main reason for this is the inadequacy of the computed outline for objects with see-through holes. That is, interior outlines delimiting the object and the holes would be needed as well, in order to better extract parts. Other reasons could be unfamiliar viewpoints or shapes, noisy CCP maps, or noisy object outlines, all of which will have a negative effect on image retrieval. For some objects, the sets of simplified parts obtained are interpreted more consistently from view to view (see Fig. 6 for examples of good and bad clusters). This is particularly so in the case of the lamp and the compass. On average, the proposed method behaves similarly to humans with respect to the ease or difficulty of recognizing object categories, although almost always with less purity. An exception is the desk lamp category. In this case, humans have no clear advantage over the proposed method, since part maps are very consistent, and neither further reasoning nor prior knowledge is of any use. The same is true for the entropy values obtained.

The main advantage human subjects have over our method is their knowledge of the global structure and precise positioning of the parts. The graph of categorized parts does not include this information. Coarse positioning is used for graph

edges, but the global shape arising from the parts is not used. Our method could be improved by adding this information to the description.

Figure 6

5.2 Image retrieval experiments

PLASTIQUE, a CBIR system, was developed and used in the image retrieval experiments. It integrates object outline detection from a CCP map, part segmentation, the 3D part categorization processes described in this paper, part spatial structure extraction, and model matching using fuzzy structural indexing.

Fig. 7A illustrates how our interpretation method should allow the matching of different images of lamps by generating many part categorization hypotheses. Fig 7B shows that in practice the proposed method integrated into PLASTIQUE indeed provides some degree of viewpoint invariance (see section 5.1). A feature of PLASTIQUE illustrated in the figure is the possibility of using a sketch instead of an image as a query.

Figure 7

Fig. 8A presents standard precision versus recall graphs for an image retrieval experiment. A database of 312 images with 11 object categories is used. The distribution of the categories is 34 watering cans, 29 airplanes, 24 chairs, 40 compasses, 39 lamps, 23 ironing boards, 27 stools, 24 coffee cups, 12 screwdrivers, 16 electric screwdrivers, and 44 desklamps. Average precision (for all instances as a query) versus recall is displayed for each of the object categories. The overall average is also displayed. For five object categories (screwdriver, compass, coffee cup, desk lamp, and watering can), the results are better than average. For two object categories (chair and ironing board), the

results are similar to average. Finally, for the last three object categories (stool, airplane, and electric screwdriver), the results are below average.

Figure 8

These results are coherent with the nature of the objects and the previous clustering experiment. Finer analysis shows that existing difficulties are not strictly caused by viewpoint and within-category shape variations, although some extreme viewpoint differences may result in poor similarity. The object categories for which the proposed method is less precise have poor segmentation, noisy object outlines, missing parts due to holes (chair, stool, ironing board), or complex CCP maps (airplane), or a combination of these (Bilodeau and Bergevin, 2002). See Figs. 5 and 6 for examples of poor stool segmentations caused by an undetected internal outline. Work is under way to improve object outline detection (Randrianarisoa et al., 2005, Bernier and Bergevin, 2006).

One might wonder if the results obtained are simply "chance" results, where images are ordered randomly. In Fig. 8B, our method is compared to results from random ordering. Random ordering has been evaluated by randomly generating results. Our method performs significantly better than random results. This is coherent with the purity results obtained for the clustering experiment, which showed that many images within a given category share the same description.

We now compare the performance of our method to methods based on low-level features. GIFT 0.1.13, an open-source implementation of Viper (Squire et al., 2000) is used as a comparison baseline. The overall average precision versus recall is displayed in Fig. 8B for the proposed method, using qualitative volumetric part-based models, and for GIFT, using a color histogram, color

blocks, and Gabor filters for textures. GIFT performs better with the present database. This was to be expected, since many images present different views of the same object instance on a uniform background. For example, in the chair category, all images display the same chair from different viewpoints. The colors and textures being the same, GIFT performs flawlessly across the recall range, while our method is handicapped by segmentation errors. Watering can images, by contrast, have different backgrounds and changing foreground colors. For that category, the performance of GIFT drops rapidly, while our method performs better. This is because, although the colors are different, the structure of the watering can is always the same, and hence using a qualitative part-based model is advantageous. We may conclude that the choice of image retrieval method depends on what information is considered important by the user. If it is the identity of an object regardless of its color that is important, then a part-based method will perform better, whereas if color is an important discriminative criterion, then low-level methods should be selected.

6. Conclusion

In this paper, the categorical invariance and viewpoint invariance of a new qualitative volumetric part-based model is assessed. A categorization method was developed to interpret the object parts extracted from an image as qualitative volumetric primitives. The proposed categorization method first transforms parts into simplified parts that correspond to possible projections of a set of chosen simple qualitative volumetric primitives. The template parts are interpreted as volumetric primitive hypotheses by rules based on geometric attributes.

Validating experiments have confirmed that, in practice, a qualitative volumetric part-based model handles viewpoint and within-category shape variations. Considering image parts as projections of volumetric primitives gives better

performance than matching them as 2D shapes. Although not as effective as human subjects, the proposed method produces a small number of different descriptions for each complex object category. Some descriptions deviate from the main descriptions due to part segmentation errors at see-through holes. This could be improved by segmenting simultaneously from contours and regions, assuming that the background seen through holes is similar to the background around the object. Our conjecture as to why our method is outperformed by human subjects is that no use is made in the proposed method of a priori knowledge or global information about the spatial distribution of the parts. Our method could be improved by integrating a more constrained relative positioning of the parts.

Further validating experiments used an image retrieval task to confirm that the proposed method does allow matching instances of given object types presenting within-category shape variations and seen from different viewpoints. This capability results from the use of many volumetric primitive hypotheses to account for different views of each object part. Using the coherent descriptions produced, the fuzzy matching method matches objects experiencing unrestricted viewpoint changes. Compared to methods using only low-level features, a model that accounts for structure can perform better on queries based on the structure or identity of objects. Therefore, image retrieval and even video surveillance could benefit from higher-level qualitative volumetric part-based models.

Acknowledgments We would like to thank Cynthia Hunt for revising the paper and the anonymous reviewers for providing helpful comments.

References

Bergevin, R., Levine, M. D., Generic Object Recognition: Building and Matching Coarse Descriptions from Line Drawings, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 1, January 1993, pp. 19-36.

Bernier, J.-F., Bergevin, R., Generic detection of multi-part objects, *18th International Conference on Pattern Recognition*, Hong Kong, China, August 2006.

Biederman, I., Human Image Understanding: Recent Research and a Theory, *Computer Vision Graphics Image Processing*, 32 (1985), pp. 29-73.

Bilodeau, G.-A., Bergevin, R., Generic modeling of 3D objects from single 2D images, *IAPR 15th International Conference on Pattern Recognition*, Barcelona, Spain, 2000, pp. 770-773.

Bilodeau, G.-A., Bergevin, R., Part segmentation of objects in real images, *Pattern Recognition*, vol. 35, No. 12, December 2002, pp. 2913-2926.

Bilodeau, G.-A., Bergevin, R., Matching graphs with fuzzy attributes in machine vision, *Acta Press International Journal on Robotics and Automation*, Vol. 20, No. 1 (2005), pp. 50-59.

Dhillon, I., Fan, J., and Guan, Y., Efficient Clustering of Very Large Document Collections, In R. Grossman, G. Kamath, and R. Naburu, editors, *Data Mining for Scientific and Engineering Applications*, Kluwer Academic Publishers, 2001, pp. 15-16.

Dickinson, S. J.., Pentland, A. P., Stevenson, S., Viewpoint-invariant indexing for content-based image retrieval, *IEEE International Workshop on Content-based Access of Image and Video Databases*, Bombay, India, 1998, pp. 20-30.

Dickinson, S. J., Pentland, A. P., Rosenfeld, A., 3-D Shape recovery Using Distributed Aspect Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, February 1992, pp. 174-198.

Dorko, G., Schmid, C., Selection of scale-invariant parts for object class recognition, *Proc. Ninth International Conference on Computer Vision (ICCV'03)*, Nice, France, 2003, pp. 20-30.

Elder, J. H., Zucker, S.W., Computing contour closure, *Proc. 4th European Conference on Computer Vision*, Cambridge, England, 1996, pp. 399-411.

Forsyth, D.A., Ioffe, S., Haddon, J., Finding objects by grouping primitives, *Conf. Rec. of 32nd Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, 1998, pp. 905-909.

Gross, A.D., Boult, T. E., Recovery of SHGCs from a Single Intensity View, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 2, February 1996, pp. 161-180.

Hérault, L., Horaud, R., Figure-Ground Discrimination: A Combinatorial Optimization Approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, September 1993, pp. 899-914.

Jacobs, D. W., Robust and Efficient Detection of Salient Convex Groups, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 1, January 1996, pp. 23-37.

Jacot-Descombes, A., Pun, T., Asynchronous Perceptual Grouping: From Contours to Relevant 2-D Structures, *Computer Vision and Image Understanding*, Vol. 66, No. 1, April 1997, pp. 1-24.

Liu, J., Mundy, J., Forsyth, D., Zisserman, A., Rothwell, C., Efficient Recognition of Rotationally symmetric surfaces and straight homogenous generalized cylinders, *Proc. IEEE Computer Vision and Pattern Recognition*, New York, NY, 1993, pp. 123-128.

Liu, W.; Zhang, C.; Yuan, B., Superquadric-based reconstruction from 2D images, *Proc of 6th International Conference on Signal Processing*, Beijing, China, August 26-30, 2002, pp. 668-671.

Martin, D., Fowlkes, C., Malik, J., Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5, May 2004, pp. 530-549.

Mokhtari, M., Bergevin R., Generic Multi-scale Segmentation and Curve Approximation Method, *Proc. LNCS 2106: Scale-Space and Morphology in Computer Vision, Third Int. Conf.*, Vancouver, Canada, 2001, pp. 227-235.

Mokhtarian, F., Silhouette-Based Object Recognition with Occlusion through Curvature Scale-Space, *Proc. 4th European Conference on Computer Vision*, Cambridge, England, 1996, pp. 566-578.

Nishida, H., Structural Feature Indexing for Retrieval of Partially Visible Shapes, *Pattern Recognition*, 35, 1, January 2002, pp. 55-67.

Pilu, M., Fisher, R. B., Recognition of Geons by Parametric Deformable Contour Models, *Proc. European Conference of Computer Vision*, Cambridge, England, in Lecture Notes in Computer Science, LNCS 1064, Springer-Verlag, Berlin. April 1996, pp. 71-92.

Ponce, J., Chelberg, D., Mann, W. B., Invariant Properties of Straight Homogenous Generalized Cylinders and Their Contours, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 9, September 1989, pp. 951-966.

Randrianarisoa, V., Bernier, J.-F., Bergevin, R., Detection of multi-part objects by top-down perceptual grouping, *Proc. 2nd Canadian Conference on Computer and Robot Vision*, Victoria, Canada, 2005, pp. 536-543.

Sato, H., Binford, T. O., Finding and Recovering SHGC Objects in an Edge Image, *CVGIP: Image Understanding*, Vol. 57, No. 3, May 1993, pp. 346-358.

Schaffalitzky, F., Zisserman, A., Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?" *Proc. 7th European Conference on Computer Vision*, Copenhagen, Denmark, 2002, pp. 414-431.

Selinger, A., Nelson, R. C., A Perceptual Grouping Hierarchy for Appearance-Based 3D Object Recognition, *Comp. Vision and Image Understanding*, Vol. 76, No. 1, October 1999, pp. 83-92.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., and Zucker, S. W., Shock Graphs and Shape Matching, *International Journal of Computer Vision*, 35(1), 1999, pp. 13-32.

Singh, S., G. D. Seyranian, and D. D. Hoffman, Parsing Silhouettes, *Perception and Psychophysics*, (61) (1999), pp. 636-660.

Sivic, J., Zisserman, A., Video Google: a text retrieval approach to object matching in videos, *Proc. Ninth IEEE International Conference on Computer Vision*, Nice, France, 2003, pp .1470-1477.

Squire, D. M., Müller, W., Müller, H., Pun, T., Content-based query of image databases: inspirations from text retrieval, *Pattern Recognition Letters*, 21, 2000, pp. 1193-1198.

Stein, F., Medioni, G., Structural Indexing: Efficient 2D Object Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 12, 1992, pp. 1198-1204.

Zerroug M., Nevatia, R., Part-based 3D Description of Complex Objects from a Single Image, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 9, 1999, pp. 835-848.

Appendix

Expressions to compute four attributes of simplified parts are described.

Straight-line segment parallelism (pSLS)

$$pSLS = \begin{cases} 2, \text{ if two pairs of } SLS \text{ have } NP \geq 0.5 \\ 1, \text{ if one pair of } SLS\text{s has } NP \geq 0.5 \\ 0, \text{ if all pairs of } SLS\text{s have } NP < 0.5 \end{cases} \quad (6)$$

with:

$$NP = \frac{1}{1 + e^{(-((1-\Delta s)-(0.75+(L/410)*(1-(P/400)))*(25+0.56*L)))}} \quad (7)$$

$$\Delta s = \frac{|d1 - d2|}{2L} \quad (8)$$

The sweeping rule variation by axis unit (Δs) is computed using the straight-line segment (*SLS)* symmetry axis of a pair of *SLSs*. The length of the axis is *L* and the sizes of sections at the extremities of the axis are *d1* and *d2* respectively.

If $|d1 - d2| > 2L$, the *SLS*s are not considered parallel, since the angle between the two *SLS*s is larger than 90°. If the *SLS*s are perfectly parallel, Δs =0. If the *SLS*s are perpendicular, Δs =1. To account for the length of the axis and the distance between the *SLS*s, a function that modifies Δs accordingly is required. A sigmoid function with a 0.5 threshold is used.

Circular arc compatibility (cCA)

$$cCA = \begin{cases} 2, \text{ if two pairs of } CA\text{s are compatible} \\ 1, \text{ if a single pair of } CA\text{s is compatible} \\ 0, \text{ if no pair of } CA\text{s is compatible} \end{cases} \quad (9)$$

Two circular arcs are compatible if their sectors overlap (Bergevin and Levine, 1993). Since the endpoints of the CCPs composing the simplified parts are ordered clockwise and circular arcs are also defined clockwise, it is easy to determine compatibility using the position of the endpoints on each circle.

Convexity of CCPs (Conv)

$$Conv = \begin{cases} AC, & \text{no } CA \text{ is convex.} \\ TC, & \text{all } CA\text{s are convex.} \\ ECOMP, & \text{if } CA\text{s between compatible } CA\text{s are convex.} \\ !ECOMP, & \text{if } CA\text{s between compatible } CA\text{s are concave.} \\ CC, & \text{if a pair of non consecutive } CA\text{s are convex} \\ & \text{and a pair of non consecutive } CCP\text{s are concave.} \end{cases} \quad (10)$$

Convexity is determined using the same principle as is used for compatibility. A *CA* is convex if its orientation in the part path is the same as in its definition. A straight-line segment is considered convex.

Sweeping rule (Sweep)

$$Sweep = \begin{cases} Constant, & \text{if } AreaTot < 1.25 * AreaConst \\ Increasing, & \text{if } AreaTot > 1.25 * AreaConst \\ & \text{and (Section size increasing or decreasing)} \\ IncreasingDecreasing, & \text{else} \end{cases} \quad (11)$$

This last attribute is used for simplified parts with compatible circular arcs. It is established by comparing the area between the compatible arcs and the area when the *Constant* rule is assumed.

Figure 1

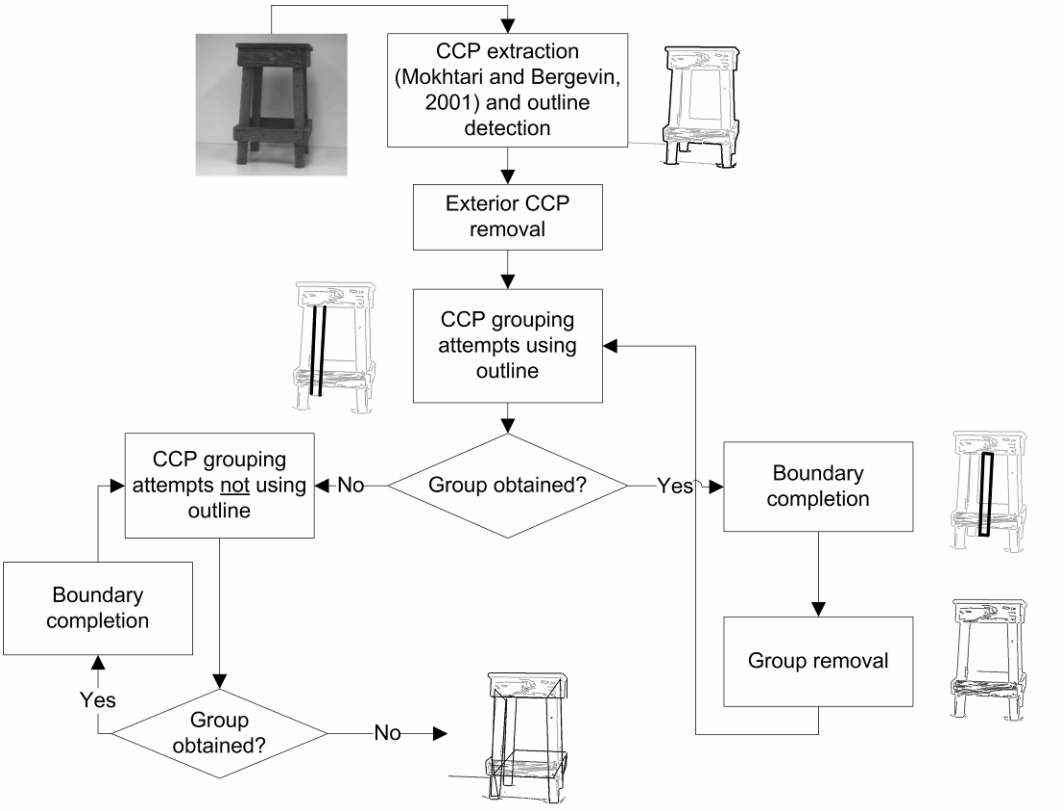

CCP extraction (Mokhtari and Bergevin, 2001) and outline detection
Exterior CCP removal
CCP grouping attempts using outline
Group obtained?
No
CCP grouping attempts not using outline
Yes
Boundary completion
Group removal
Boundary completion
Yes
Group obtained?
No

Figure 2

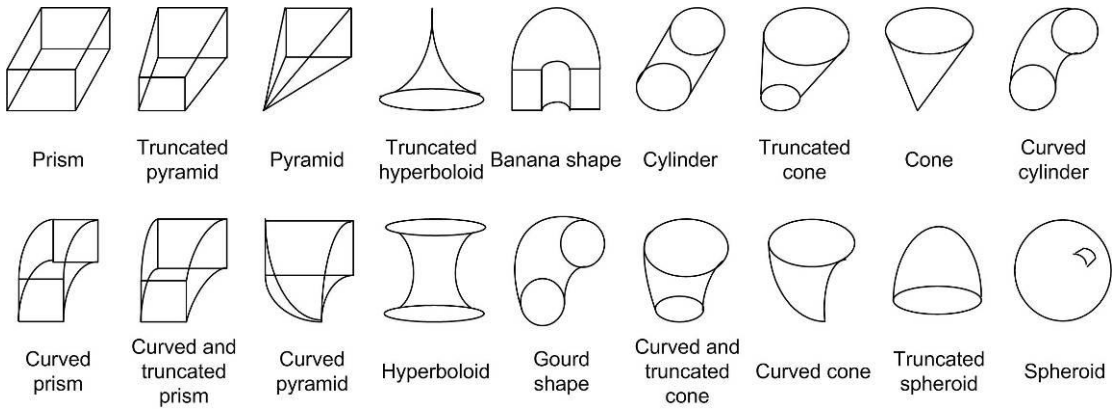

Prism
Truncated pyramid
Pyramid
Truncated hyperboloid
Banana shape
Cylinder
Truncated cone
Cone
Curved cylinder
Curved prism
Curved and truncated prism
Curved pyramid
Hyperboloid
Gourd shape
Curved and truncated cone
Curved cone
Truncated spheroid
Spheroid

Figure 3

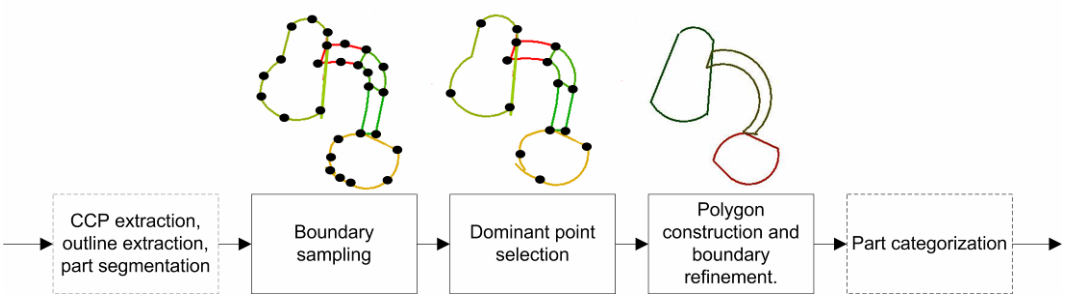

CCP extraction, outline extraction, part segmentation
Boundary sampling
Dominant point selection
Polygon construction and boundary refinement.
Part categorization

Figure 4

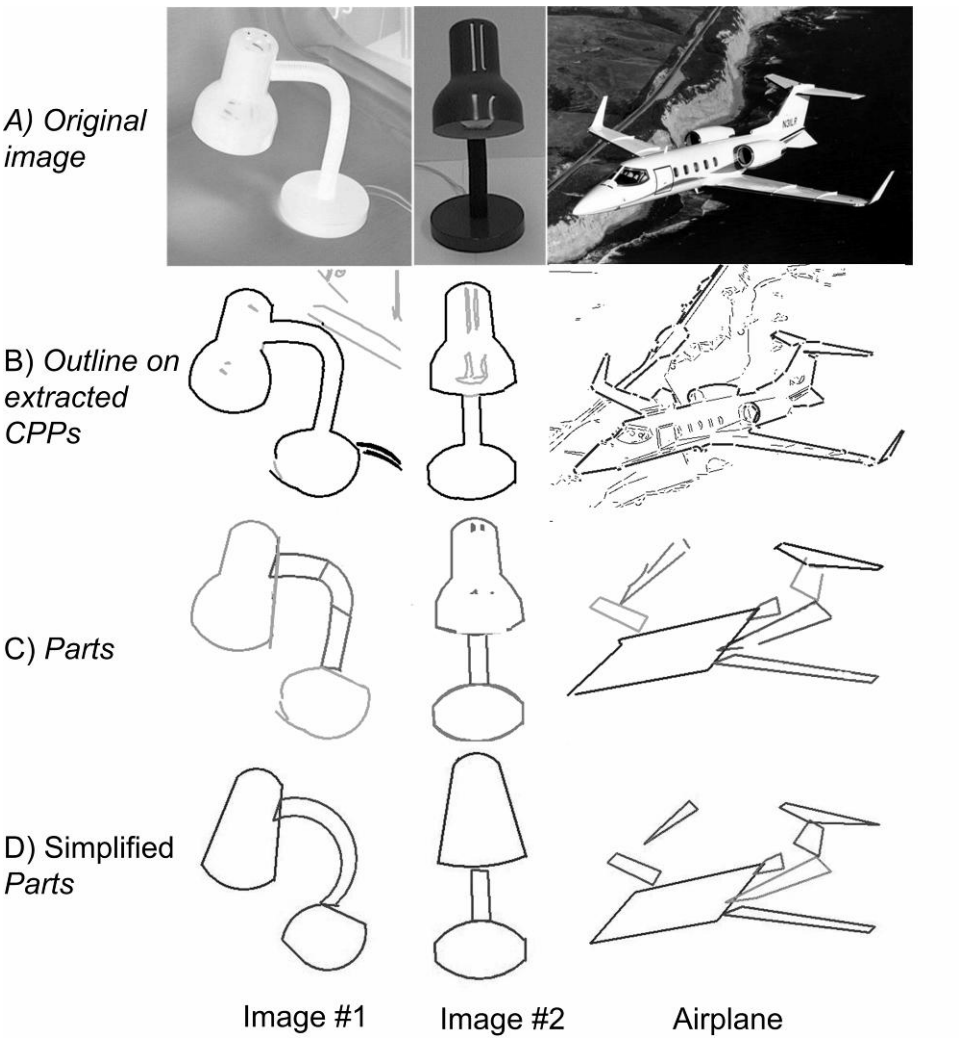

A) Original image
B) Outline on extracted CPPs
C) Parts
D) Simplified Parts
Image #1
Image #2
Airplane

Figure 5

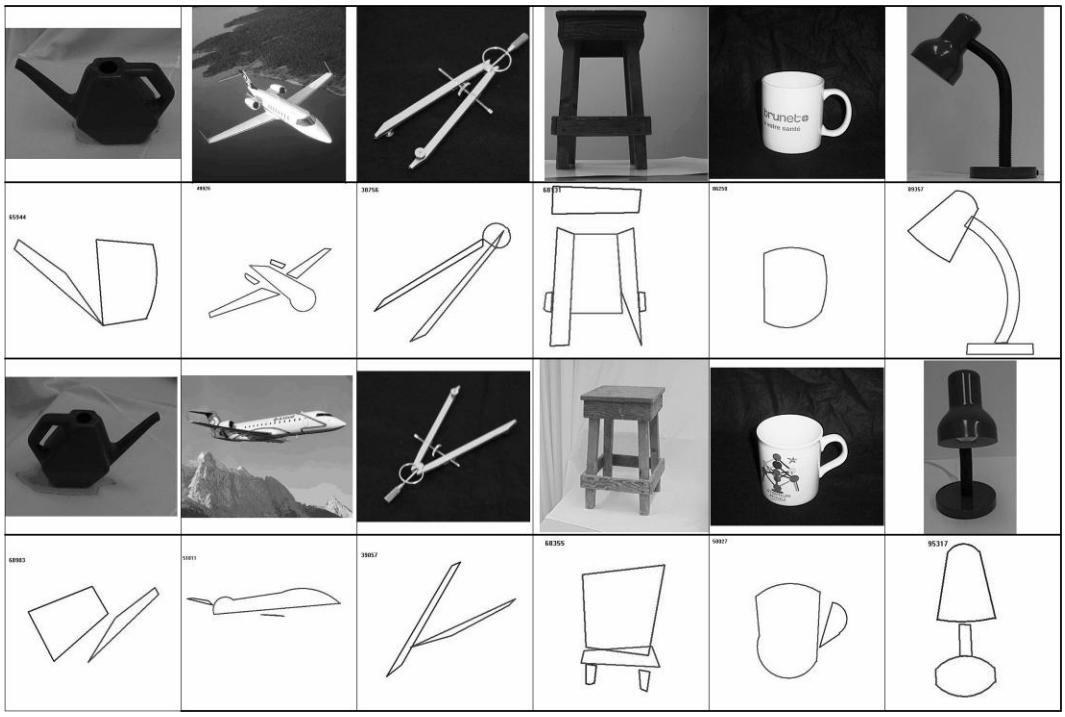


Figure 6

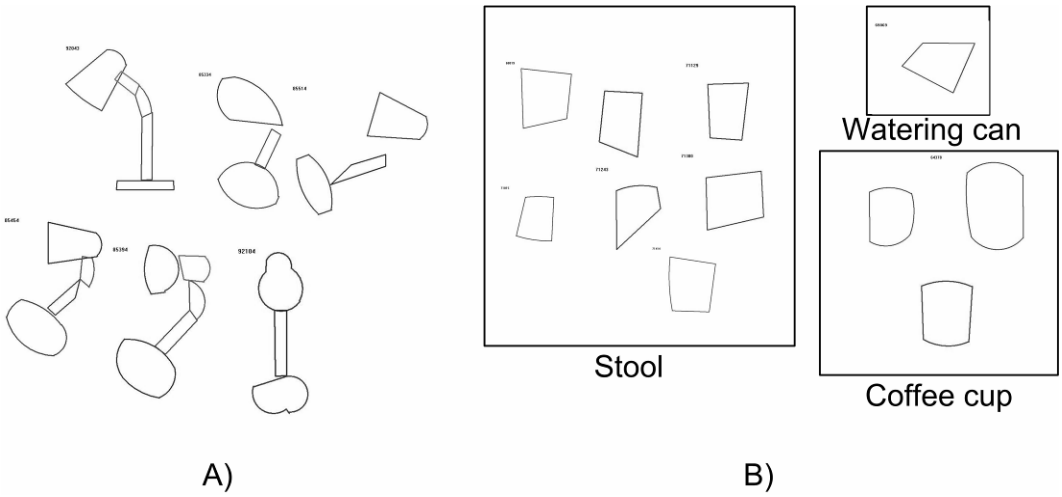
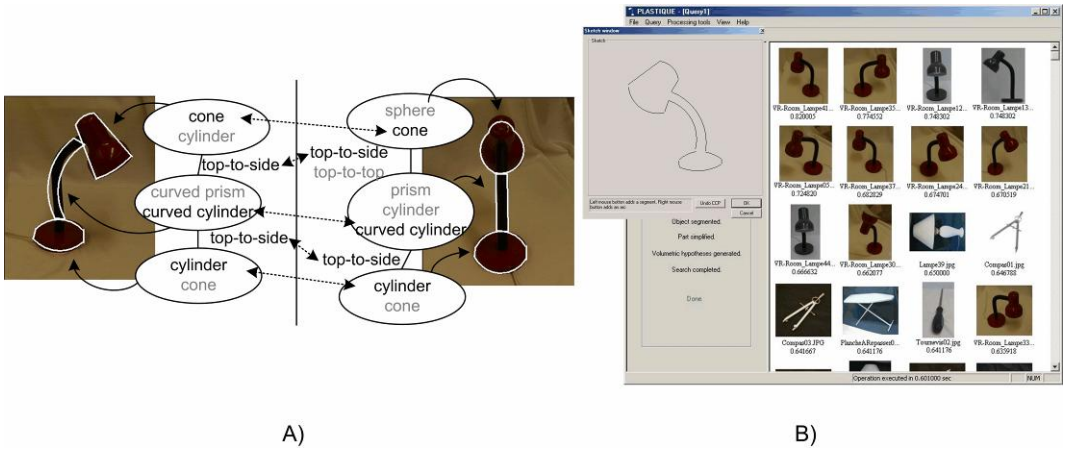
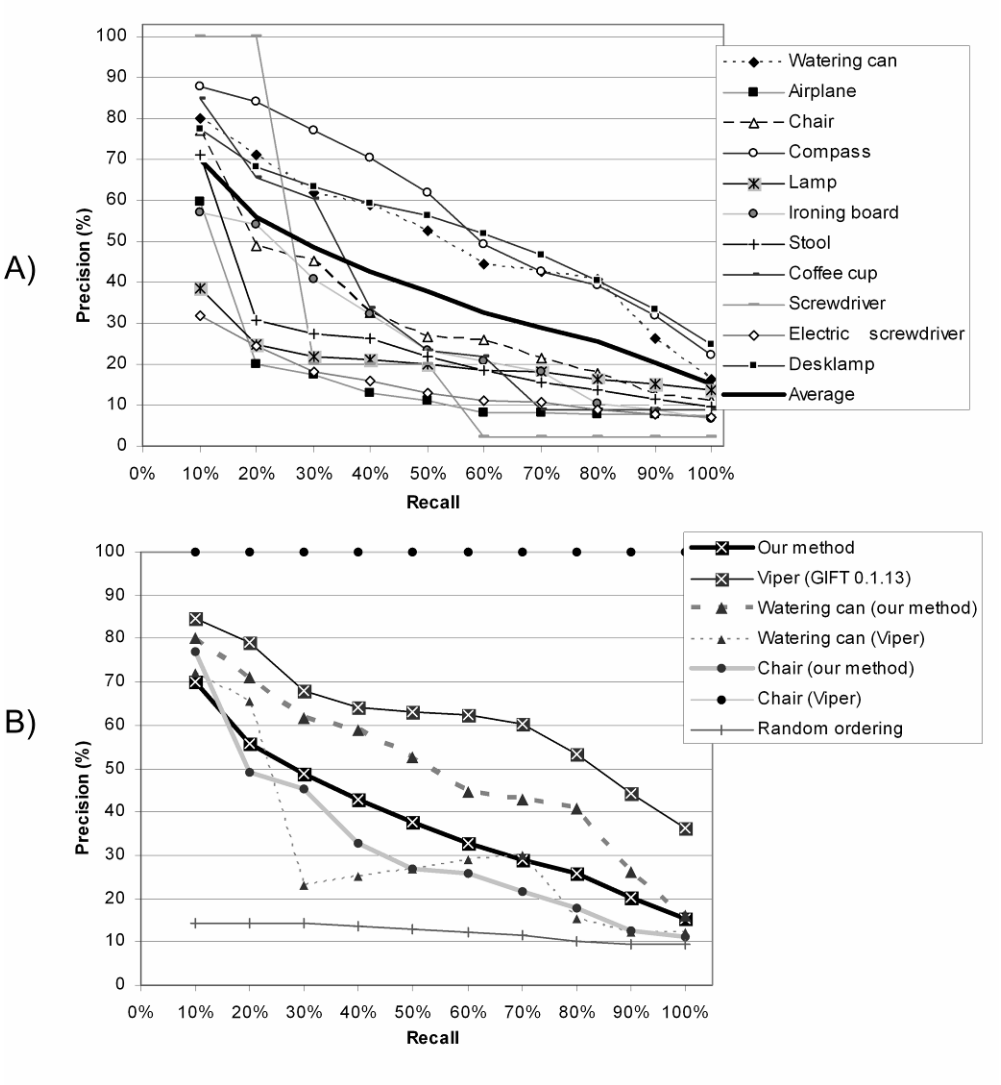


Figure 7


cone
cylinder
top-to-side
curved prism
curved cylinder
top-to-side
cylinder
cone
sphere
cone
top-to-side
top-to-top
prism
cylinder
curved cylinder
top-to-side
cylinder
cone
PLASTIQUE - [Query1]
File Query Processing tools View Help
Object segmented.
Part simplified.
Volumetric hypotheses generated.
Search completed.
Done.

A) B)

Figure 8


A)
Precision (%)
Recall
Watering can
Airplane
Chair
Compass
Lamp
Ironing board
Stool
Coffee cup
Screwdriver
Electric screwdriver
Desklamp
Average
B)
Precision (%)
Recall
Our method
Viper (GIFT 0.1.13)
Watering can (our method)
Watering can (Viper)
Chair (our method)
Chair (Viper)
Random ordering

Legend

Fig. 1 Flow diagram of part segmentation with illustrative images.

Fig. 2 The eighteen qualitative volumetric primitives used in the proposed method.

Fig. 3 Flow diagram of part transformation with illustrative images.

Fig. 4 Processing step results from input image to simplified parts.

Fig. 5 Typical printouts of simplified parts used in a clustering experiment with their corresponding image.

Fig. 6 Cluster examples. A) A good cluster of a desk lamp; B) A bad cluster, mixing stools, watering cans, and coffee cups.

Fig. 7 Viewpoint invariance. A) In theory; B) In practice, with a result from a sketch query. Retrieved database images are sorted according to decreasing similarity.

Fig. 8 A) Precision versus recall graph with PLASTIQUE; B) Comparison with Viper (GIFT 0.1.13) and random ordering.

Table 1. Classifier rules

#	Part	Premise	Conclusion (ranking value)
1		n*CCP*s=4 & n*CA*s=0 & p*SLS*=2	Prism(1), t. pyr. (0.9), cyl. (0.8), t. cone (0.7), pyr. (0.4)
2		n*CCP*s=4 & n*CA*s=0 & p*SLS*=1	T. pyr. (1), prism (0.9), t. cone (0.8), cyl. (0.7), pyr. (0.4)
3		n*CCP*s=4 & n*CA*s=0 & p*SLS*=0	T. pyr. (1), prism (0.9), t. cone (0.8), cyl. (0.7), pyr. (0.4)
4		nCCPs=3 & nCAs=0	Pyr. (1), cone (0.9), T. pyr. (0.4)
5		nCCPs=3 & nCAs=2 & Conv.=AC	T. hyp. (1)
6		nCCPs=3 & nCAs=2 & cCA=1	C. pyr. (1), c. cone (0.9)
7		nCCPs=3 & nCAs=2 & Conv=TC	T. sph. (1)
8		nCCPs=4 & nCAs=2 & pSLS=1 & Conv=TC	Cyl. (1), t. cone (0.9), t. sph. (0.8)
9		nCCPs=4 & nCAs=1 & pSLS=1 Conv=TC	Cyl. (1), t. cone (0.9), c. cyl. (0.4), c. prism (0.4)
10		nCCPs=4 & nCAs=2 & pSLS=0 & Conv=TC	T. cone (1), cyl. (0.9), T. sph. (0.8)
11		nCCPs=4 & nCAs=1 & pSLS=0 & Conv=TC	T. cone (1), cyl. (0.9), c. cyl. (0.4), c. prism (0.4)
12		nCCPs=4 & nCAs=2 & pSLS=1 & cCA=1 & sweep=inc.	T.c. pyr. (1), t.c. cone (0.9), c. prism (0.8), c. cyl. (0.7), cyl. (0.4)
13		nCCPs=4 & nCAs=2 & pSLS=0 & cCA=1 & sweep=inc.	T.c. pyr. (1), t.c. cone (0.9), c. prism (0.8), c. cyl. (0.7), t. cone (0.4)
14		nCCPs=4 & nCAs=2 & pSLS=1 & cCA=1 & sweep=inc.&dec.	Ban. (1), gou. (0.9), c. prism (0.8), c. cyl. (0.7), cyl. (0.4)
15		nCCPs=4 & nCAs=2 & pSLS=0 & cCA=1 & sweep=inc.&dec.	Ban. (1), gou. (0.9), c. prism (0.8), c. cyl. (0.7), cone (0.4)
16		nCCPs=4 & nCAs=2 & Conve=AC & pSLS=1	Hyp. (1), cyl. (0.4)
17		nCCPs=4 & nCAs=1 & Conv=AC & pSLS=1	Cyl. (1), c. prism (0.4), c. cyl. (0.4)
18		nCCPs=4 & nCAs=2 & Conv=AC & pSLS=0	Hyp. (1), cone (0.4)
19		nCCPs=4 & nCAs=1 & Conv=AC & pSLS=0	Cone (1), c. prism (0.4), c. cyl. (0.4)
20		nCCPs=4 & nCAs=2 & pSLS=1 & cCA=1 & sweep=const.	C. prism (1), c. cyl. (0.9), ban. (0.8), t.c. pyr. (0.75), gou. (0.7), t.c. cone (0.65), cyl. (0.4)
21		nCCPs=4 & nCAs=2 & pSLS=0 &cCA=1 & sweep=const.	C. prism (1), c. cyl. (0.9), ban. (0.8), t.c. pyr. (0.75), gou. (0.7), t.c. cone (0.65), cone (0.4)
22		nCCPs=3 & nCAs=1 & Conv=TC	Cone (1), c. prism (0.4), c. cyl. (0.4), t. sph. (0.4)
23		nCCPs=3 & nCAs=1 & Conv=AC	Cone (1)
24		nCCPs=3 & nCAs=3 & cCA=1 & Conv=ECOMP	C. cone (1), c. cyl. (0.4)
25		nCCPs=3 & nCAs=3 & cCA=1 & Conv=!ECOMP	T. hyp. (1), c. cone (0.4)
26		nCCPs=4 & nCAs=4 & Conv=TC	Sph. (1), cyl. (0.9), cone (0.8), t. sph. (0.7)
27		nCCPs=4 & nCAs=3 & Conv=TC	T. sph. (1), sph. (0.8), cyl. (0.7), cone (0.6)
28		nCCPs=4 & nCAs=4 & cCA=1 & Conv=ECOMP & sweep=const.	C. cyl. (1), gou. (0.9), t.c. cone (0.85), t. sph. (0.4)
29		nCCPs=4 & nCAs=3 & cCA=1 & Conv=ECOMP & sweep=const.	C. cyl. (1), gou. (0.9), t.c. cone (0.8)
30		nCCPs=4 & nCAs=4 & cCA=1 & Conv=!ECOMP & sweep=const.	C. cyl. (1), gou. (0.9), hyp. (0.8), t. cone (0.8)
31		nCCPs=4 & nCAs=3 & cCA=1 & Convexe=!ECOMP & Balayage=Constant	C. cyl. (1), gou. (0.9), t.c. cone (0.8)
32		nCCPs=4 & nCAs=4 & cCA=0 &Conv=CC	Hyp. (1), t. sph. (0.4)
33		nCCPs=4 & nCAs=3 & cCA=0 &Conve=CC	Hyp. (1)
34		nCCPs=4 & nCAs=3 & cCA=0 &Conv=CC	Cyl. (1), t.sph (0.9)
35		nCCPs=4 & nCAs=4 & cCA=0 & Conv=AC	Hyp. (1)
36		nCCPs=4 & nCAs=3 & cCA=0 & Conv=AC	Hyp. (1), cyl. (0.4)
37		nCCPs=4 & nCAs=4 & cCA=1 & Conv=ECOMP & sweep=inc.&dec.	Gou. (1), c. cyl. (0.9), t. sph. (0.8)
38		nCCPs=4 & nCAs=3 & cCA=1 & Conv=ECOMP & sweep=inc.&dec.	Gou. (1), c. cyl. (0.9)
39		nCCPs=4 & nCAs=4 & cCA=1 & Conv=!ECOMP & sweep=inc.&dec.	Gou. (1), c. cyl. (0.9), hyp. (0.8)
40		nCCPs=4 & nCAs=3 & cCA=1 & Conv=!ECOMP & sweep=inc.&dec.	Gou. (1), c. cyl. (0.9)
41		nCCPs=3 & nCAs=3 & Conv=TC	Sph. (1), cyl. (0.9), cone (0.8)
42		nCCPs=3 & nCAs=3 & Conv=AC	T. hyp. (1)
43		nCCPs=4 & nCAs=4 & cCA=1 & Conv=ECOMP & sweep=inc.	T.c. cone (1), c. cyl. (0.9), t. sph. (0.4)
44		nCCPs=4 & nCAs=3 & cCA=1 & Conv=ECOMP & sweep=inc.	T.c. cone (1), c. cyl. (0.9)
45		nCCPs=4 & nCAs=4 & cCA=1 & Conv=!ECOMP & sweep=inc.	T.c. cone (1), c. cyl. (0.9), hyp. (0.8)
46		nCCPs=4 & nCAs=3 & cCA=1 & Conve=!ECOMP & sweep=inc.	T.c. cone (1), c. cyl. (0.9)
47		nCCPs=4 & nCAs=4 & cCA=2 &sweep=const.	C. cyl. (1), gou. (0.9), t.c. cone (0.8)
48		nCCPs=4 & nCAs=4 & cCA=2 & sweep=const.	T.c. cone (1), c. cyl. (0.9)
49		nCCPs=4 & nCAs=4 & cCA=2 & sweep=inc.&dec.	Gou. (1), c. cyl. (0.9)

c.=curved, t.c.=truncated and curved, t.=truncated, cyl.=cylinder, pyr.=pyramid, ban.=banana shape , gou.=gourd shape, hyp.=hyperboloid, sph.=spheroid

Table 2. Clustering experiment results

Dominated clusters											
	S1	S2	S3	S4	S5	S6	Subject average	Proposed method 70%	Proposed method 60%	Shape only 70%	Shape only 60%
Watering can	6	5	4	1	1	3	**3.33**	*3*	2	11	3
Airplane	5	4	5	1	2	4	**3.50**	*4*	2	3	3
Compass	4	1	2	1	1	3	**2.00**	*8*	7	10	6
Stool	5	6	3	2	4	4	**4.00**	*4*	3	15	3
Coffee cup	2	1	3	0	3	2	**1.83**	*3*	1	8	3
Desk lamp	8	1	2	1	1	1	**2.33**	*7*	4	5	9
Printout images not clustered											
Watering can	1	3	5	0	0	5	**2.33**	*3*	1	6	1
Airplane	8	11	13	3	0	3	**6.33**	*12*	5	26	15
Compass	1	2	4	1	1	1	**1.67**	*5*	2	7	1
Stool	4	4	6	1	0	1	**2.67**	*4*	2	5	7
Coffee cup	4	5	5	0	1	0	**2.50**	*4*	4	9	6
Desk lamp	4	3	6	1	0	1	**2.50**	*8*	4	29	7
Total	22	28	39	6	2	11	**18.00**	*36*	18	82	37
Purity											
Watering can	0.89	0.96	0.63	1.00	0.94	1.00	**0.90**	*0.77*	0.60	0.62	0.48
Airplane	1.00	0.86	0.94	0.84	0.89	0.96	**0.92**	*0.89*	0.49	0.67	0.60
Compass	1.00	1.00	1.00	1.00	0.90	0.95	**0.98**	*0.85*	0.58	1	0.54
Stool	0.78	0.68	0.55	0.41	0.79	0.68	**0.65**	*0.47*	0.41	0.61	0.53
Coffee cup	0.88	1.00	1.00		0.67	0.79	**0.87**	*0.76*	0.25	0.50	0.43
Desk lamp	1.00	1.00	1.00	0.93	1.00	1.00	**0.99**	*0.98*	0.76	0.84	0.70
Entropy											
Watering can	0.24	0.09	0.45	0.00	0.16	0.00	**0.16**	*0.41*	0.64	0.42	0.76
Airplane	0.00	0.28	0.12	0.31	0.23	0.09	**0.17**	*0.19*	0.65	0.26	0.45
Compass	0.00	0.00	0.00	0.00	0.24	0.11	**0.06**	*0.28*	0.71	0	0.60
Stool	0.49	0.49	0.61	0.71	0.44	0.50	**0.54**	*0.79*	0.77	0.32	0.81
Coffee cup	0.21	0.00	0.00		0.61	0.42	**0.25**	*0.39*	0.97	0.39	0.66
Desk lamp	0.00	0.00	0.00	0.12	0.00	0.00	**0.02**	*0.05*	0.43	0.17	0.29
S# stands for subject #											