

PLASTIQUE: Image retrieval based on cognitive theories

Guillaume-Alexandre Bilodeau and Robert Bergevin

Laboratoire de vision et systèmes numériques, Pavillon Adrien-Pouliot
Université Laval, Sainte-Foy (QC), Canada, G1K 7P4

bilodeau@gel.ulaval.ca, bergevin@gel.ulaval.ca

Abstract

This paper presents the image retrieval software PLASTIQUE. It distinguishes itself by the use of a 3D part-based model to represent objects in images. This model is based on results from studies in cognitive psychology. This model allows PLASTIQUE to match images of objects that are deformed or seen from different viewpoints. The model also abstracts colours and textures to focus on object structure. Our model and our method to build it can also be used for object recognition applications. The complete system is presented by an overview of all the processing stages. Experiments assess the performance of our model and of the modeling method used in the image retrieval application PLASTIQUE.

1 Introduction

Finding images in databases using another image as the query is currently an important research topic [1]. The idea of using an image as a query comes from the fact that it can be difficult to describe an image with words. In addition, many databases are composed of images that are not annotated by keywords. Hence, the content of the image itself has to be compared directly. Making a model of an image is not an easy problem. In the majority of efforts in this topic ([2],[3],[4],[5]), images are modeled by their colours, textures and 2D shapes. However, these models are sensitive to changes of viewpoint and changes in the textures and colours. For example, a blue desk lamp and a red desk lamp viewed from significantly different viewpoints cannot be matched. For this reason, we propose a 3D part-based model that abstracts colours and textures. This model applies to images of manufactured objects that are seen in the foreground. A method for obtaining this model from an image is proposed. This method models objects in images by graphs of volumetric primitives. The same method can be applied to model objects for object recognition applications.

This paper presents the integration, in an image retrieval software, of the different modules developed to implement our method. The main contribution of this paper is the evaluation in a whole of the latest contributions in part-based modeling of objects made by the authors ([6],[7],[8],[9]), and the assessment of their performances by an image retrieval application. Hence, it is not claimed here that our part-based modeling method outperforms other image retrieval modeling techniques. This paper essentially assesses, based on experimental results, how part-based modeling can perform in image retrieval applications.

Our review of the image retrieval literature has shown that part-based models have not yet been used in this field, because object segmentation algorithms are said to be too brittle for automation [1]. Hence, the use of a part-based object representation in image retrieval is an important contribution in itself.

The advantage of using a part-based representation in image retrieval is that objects in images, with different colors and textures, seen from different viewpoints can be matched. This is not possible with 2D shape-based, or colors and textures-based image retrieval.

Section 2 explains the basic principles behind our model and our method. Section 3 gives an overview of the method, and of PLASTIQUE, our image retrieval software. Section 4 illustrates the performance of our method by query experiments. Section 5 concludes the paper.

2 The theories behind PLASTIQUE

The representation model of objects in images is what distinguishes PLASTIQUE from other image retrieval engines. The model chosen for representing objects in images is a graph of volumetric primitives. This choice comes from the RBC (recognition-by-components) theory and by researches in perceptual grouping. The RBC theory has given us the inspiration for the model, and perceptual grouping tools to obtain the model.

2.1 RBC (Recognition-by-components)

The RBC theory [10] claims that humans recognize objects by the structure of their constitutive parts. For example, humans recognize a table because they see four legs connected to a board. Hence, objects can be modeled as graphs of volumetric primitives. Whether it is true or not that this is how humans recognize objects, this model is interesting because recognition of objects seen from different viewpoints can be accomplished by only a few images. This is why this model has been chosen. In a database, there might be only one copy of a graph of an object. If the query is a similar object, the query engine must be able to compare the graph of these two images. This model allows the query engine to compare the two objects as graphs of volumetric primitives because they can each be constructed from a single image. View-based models cannot be used, as not enough views of the object are available.

2.2 Perceptual grouping

The chosen model has to be constructed from images of objects. Perceptual grouping [11] allows us to carry out this task. Researches in cognitive psychology have pointed out that when a human is shown and asked to make groups from sets of straight line segments, they are grouped in specific ways. The straight line segments are grouped by their level of symmetry, by proximity, by similarity, and by their level of overlap. Hence, humans are sensitive to the structure in the images they see. As image processing is concerned, these geometric relationships can be used to group (or structure) straight line segments and circular arcs into boundaries of areas that can be interpreted as being projections in a plane of volumetric primitives. These areas are named parts. A graph of volumetric primitives can be constructed by hypothesizing the volumetric primitives that might have given rise to the parts found by the grouping operation.

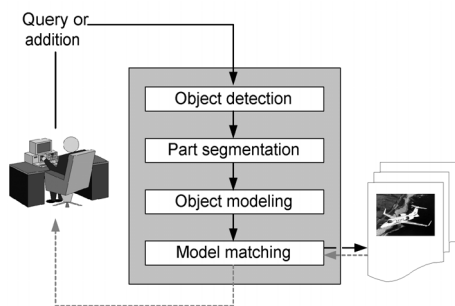


Fig. 1. Overview of PLASTIQUE

3 PLASTIQUE

PLASTIQUE (Parts, Links, and ASSociated Templates Image QUery Engine) has been designed for querying image databases of manufactured objects, which are interpretable as arrangements of simple volumetric primitives. Images in the database are from real scenes with one main object in the foreground that must be detectable in the image (see [6]).

Fig. 1 shows an overview of PLASTIQUE. The shaded region represents the four modules required to add an image or query the database. The database is composed of various 2D images of 3D objects, and their associated models. To query or add an image in the database, the user gives as input an example 2D image or a sketch of the 3D object. The image is first processed to obtain contours of linked local intensity edges that are segmented to produce a map of constant curvature primitives (CCPs). An initial grouping of the CCPs produces the outline of the object (Object detection) [6]. The CCP map is then processed further to obtain parts using the extracted outline (Part segmentation) [7]. These parts are labelled based on the possible volumetric primitives that may project onto them (Object modeling) [8]. Parts are interpreted as volumetric primitives since the aspect of a projected 3D object may change significantly for different viewpoints. The object modeling module also computes the spatial relationships between parts. Finally, the constructed model is compared with the models in the database (Model matching). If similar models are in the database, the corresponding 2D images are shown to the user. If not, the newly built model and its corresponding image may be added in the database.

The next subsections outline each of these steps and introduce PLASTIQUE graphical user interface.

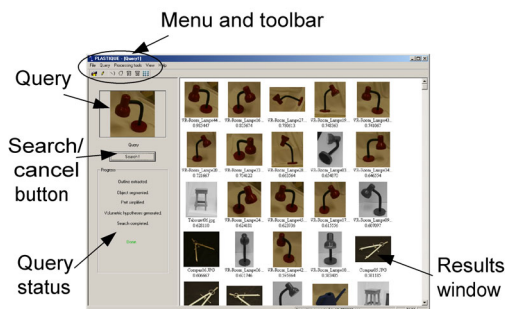


Fig. 2. PLASTIQUE graphical user interface.

3.1 PLASTIQUE graphical user interface

The graphical user interface (GUI) of PLASTIQUE is shown in Fig. 2. This GUI allows users to query and add images in the database. Images and sketches of objects can be used as query. The PLASTIQUE user interface has

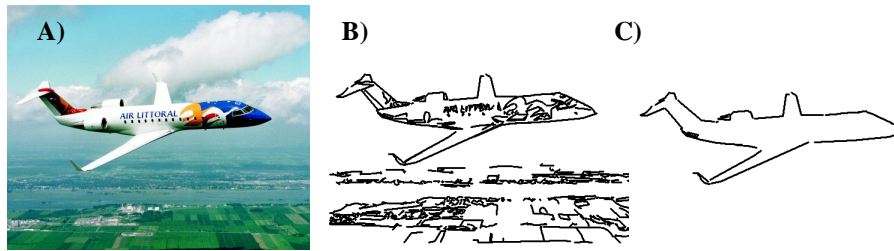


Fig. 3. A) Original airplane image, B) Extracted CCPs, C) Extracted outline.

a result window that shows query results as a sorted list of matching images. In addition, this window allows users to visualize intermediate results from the computation of the model for the object in the query image. Being a multi-documents application, PLASTIQUE can keep results of different queries in separate documents.

The PLASTIQUE user interface and PLASTIQUE query engine are distributed applications. The PLASTIQUE user interface connects to the query engine via Internet. This gives PLASTIQUE flexibility.

3.2 Extraction of low-level data and object detection

This step is essentially for information gathering. This step is used to extract the data needed for the application of perceptual grouping at the part segmentation step. First of all, an edge detector is applied on the image to process. Then, edges are grouped into edge contours. Constant curvature primitives (CCPs) are fitted on the edge contours. Edge detection, edge grouping, and CCPs extraction is performed using the algorithms of MAGNO [9]. The extracted CCPs are straight line segments and circular arcs.

To obtain only parts that are inside the object area, the external outline of the object is extracted [6]. The external outline of the object is detected by making a closed cycle of CCPs that maximizes the area to perimeter ratio. During the construction of the cycle, CCPs are added following simple rules that improve robustness to object textures. These rules involve proximity and angular distance. Fig. 3 shows the CCPs and the outline obtained for an airplane image.



Fig. 4. Parts extracted from the airplane image. The CCPs inside each part are also shown.

3.3 Part segmentation

Part segmentation [7] is the step where the construction of the models begins. Part segmentation determines the number of volumetric primitives the model will be made of, the relationships between them and the shapes of the projections. Hence, this step is critical for the quality of the model.

Starting with the longest CCPs, part segmentation groups CCPs with a geometric criterion. The components of the geometric criterion are measures of parallelism, proximity, similarity in type, similarity in length and overlap. The geometric criterion is applied on pairs of CCPs. The pair that obtains the highest values for the measures of the geometric criterion is considered as defining the main shape of a part. Joining the CCPs of the pair with CCPs nearby completes the boundary of the part. CCPs used for the part and CCPs inside the part are removed. Other part construction attempts are performed with the remaining CCPs. Several validation tests are applied on the pairs of CCPs to ensure that valid parts are obtained. A part may be invalid if it covers a region outside the extracted external outline.

When all the possible parts are obtained, their relationships are established using proximity and the sequence in which they are appearing as the external contour of the object is scanned. Two consecutive parts on the outline are linked together. Fig. 4 shows the five parts obtained for an airplane image. The complete area of the projection of an object may not be completely covered by the parts, since there is no criterion specified to optimize coverage.

3.4 Object modeling part 1: Part simplification

Object modeling can be separated into three distinct substeps. The first substep is part simplification [8]. As the name of this substep implies, its goal is to simplify parts so that a rule-based classifier can hypothesize volumetric primitives from them. The original parts can have a wide range of different shapes and they can have a boundary composed of various numbers of CCPs. Hence, a rule-based classifier cannot process these parts. An undetermined number of rules would be required to process all

the possible parts. Furthermore, we want to keep the information conveyed by the CCPs types (straight line segment or circular arcs) of the boundary. Template matching of volumetric primitives may approximate two straight line segments by a circular arc. This is why we have rejected this method. Instead, we have chosen to approximate the parts by template parts, which are all the possible qualitative projections of a set of general volumetric primitives. These template parts preserve CCPs types. Since, the number of template parts is known, a rule-based classifier can be used to hypothesize volumetric primitives with a reasonable number of rules.

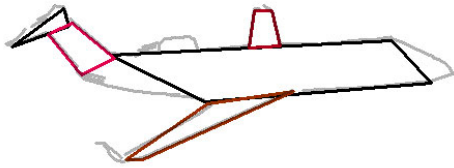


Fig. 5. Simplified parts for the airplane image.

The simplification of the parts has been designed to preserve the general shape of the original parts. Points are sampled on the boundary of a part. Among these, 3 or 4 points that maximize area coverage and rectangularity (since parts have their shape mainly defined by two CCPs) are selected. The CCPs of the boundary of the original part between these points are analysed. From this analysis, CCPs, that best approximate the original CCPs of the boundary are created between the points. The new CCPs approximate the boundary in such a way that the CCPs types are preserved as much as possible. Fig. 5 shows the simplified parts for the airplane.

3.5 Object modeling part 2: Volumetric primitive hypotheses

Volumetric primitive hypotheses [8] are performed from the simplified parts using a rule-based classifier. The rules involve geometric relationships between the CCPs of the simplified part. These relationships are parallelism, convexity, CCPs type, and the number of CCPs of the simplified part. In all, for 18 volumetric primitives [8], the classifier needs 49 rules. Each rule is associated with one or more volumetric primitive hypotheses with a fuzzy value reflecting a probabilistic ranking between the hypotheses. These hypotheses are volumetric primitives that project in the plane with shapes similar to the simplified part respecting the given rule.

3.6 Object modeling part 3: Model construction

Model construction consists in building a graph with the volumetric primitive hypotheses as nodes and their spatial relationships as edges. The constructed graph has nodes with fuzzy attributes (a volumetric hypotheses associated to a fuzzy ranking value) and has edges with fuzzy attributes (the type of connection with a probabilistic fuzzy ranking value).

3.7 Model matching

Model matching [12] is performed using structural indexing adapted to graphs with fuzzy attributes. Graphs of images in the database are inserted into an index where each row corresponds to a subgraph of 1, 2 or 3 nodes. Therefore, before adding graphs in the index, they are divided into a subgraph of 1, 2, and 3 nodes. When the database is searched, subgraphs of the query are indexed. Each time a subgraph is indexed, all the images in the database that have this subgraph receive a vote. The images that obtain the most votes are the ones that have graphs that most resemble the query image graph.

4 Experiments

To validate our model and our model construction method, PLASTIQUE has been tested for querying images. The database used has 135 images of six objects: an airplane, a stool, a lamp, a coffee cup, a compass and a watering can. Sample images of these objects and simplified parts obtained are shown in Fig. 6. Before statistically measuring the performance of PLASTIQUE to query images, we show a query result to illustrate the capabilities of our method.

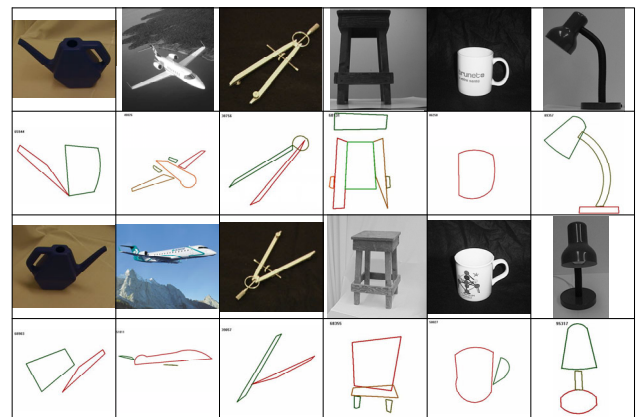


Fig. 6. Sample object images and associated simplified parts used in the experiments.

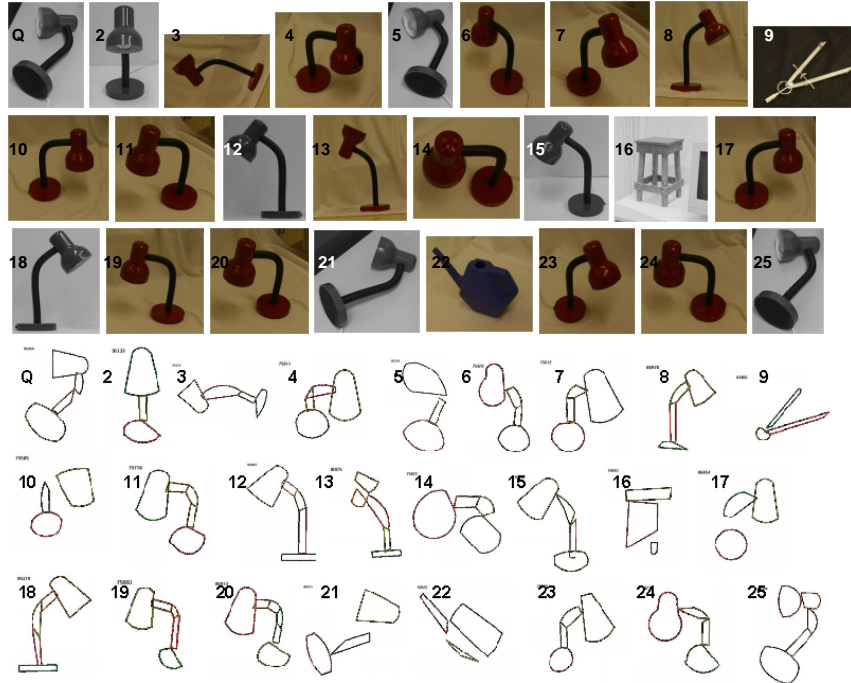


Fig. 7. Result from the query of a lamp image (image Q). The top of the figure shows the result as images. The bottom of the figure shows the result as simplified parts images. The numbers indicate the ranking between the images in terms of similarity to the query.

4.1 Example of a matching result

Fig. 7 shows a query result for a lamp image. The 25 most similar images are shown. As it should, the query, which is in the database, ranked first. The rank reflects the level of similarity of the model of an image in the database with the model of the query image. From this result, several important observations can be made. The model of the query lamp image is matched with other models of images where the lamp is seen from significantly different viewpoints. In particular, this is the case for image 2, 6 and 14. The model of the query lamp is matched also with models of images of deformed lamps. This is case for image 3, 4 and 13. The deformations do not affect the model of the lamp because the same volumetric primitives can still represent the deformed pole. These two observations demonstrate the advantages of modeling objects by volumetric primitives.

There are some models of images of other objects that are considered resembling the model of the query lamp. The graphs of these images have nodes and groups of nodes similar to the query graph. Since our matching method is not exact and the graphs are not compared as a whole, these graphs are considered quite similar. For example, they might be similar when compared node by node and by groups of two nodes, but different by groups of three nodes. Because there are variations in the models obtained for each lamp image, some models of lamps obtain a matching score that is inferior to images of other

objects (images 21 and 25, for example). In this case, it can be observed that the model of the query lamp have 4 parts, while model of image 21 has three parts and the model of image 25 has five parts. If all lamps had the same model graph, this phenomenon would not occur. However, starting from images of a real scene, as is the case here, it is difficult to always obtain the same model.

In any case, this result shows, based on experimental results, advantages and drawbacks of using a part-based representation. Although the result shown in Fig. 7 is not perfect, it shows that part-based models are useful in certain circumstances. This is the case, for deformed objects and objects seen in different viewpoints. The next section investigates furthermore the performance of our model.

4.2 Precision vs recall

The previous section has shown qualitatively the abilities of our method. This section analyses our method by quantitative measures. To verify how our method performs, we have chosen to measure its precision for different recall values. This measure reflects how many images had to be found to get $x\%$ of the images of object y . This is the standard definition of "precision vs recall". The precision in percentage is formally,

$$\text{precision} = \frac{(\%recall) * N_{image(i)}}{F} * 100\% \quad (1)$$

where $N_{image}(i)$ is the number of images that feature object i , F is the minimum number of images needed to find the $(\%recall) * N_{image}(i)$ images of object i in the database, and $\%recall$ is the proportion of the number of images of a given object to find.

The experiment to measure the “precision vs recall” values of PLASTIQUE has been conducted as follows. First, a database of 135 images of six objects has been constructed. The objects are the same as the one mentioned in the previous section. Each of the 135 images was used as a query. For each query result, the “precision vs recall” values have been computed from the sorted list (based on similarity ranking) of images obtained. Precision values have been computed for recall of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%. This experiment has been repeated for different sets of criteria. Model matching can be done with no additional criteria, as they are not essential. That is the proportions and relative dimensions are ignored. However, it was of interest for us to verify if the use of additional information could improve matching performance. We have implemented three criteria that can be applied separately or in combination. The first criterion concerns the proportion of the volumetric primitives. With this criterion, volumetric primitives (single graph nodes) can be matched only if their axis to section ratio is similar qualitatively. The second criterion applies to subgraphs of two nodes. Subgraphs of two nodes can be matched if the relative dimension of the volumetric primitives of each node is the same. The third criterion is identical to the second, but applies to subgraphs of three nodes.

Fig. 8 shows “precision vs recall” plots when matching without criteria and with criteria 1 and 2. When matching without criteria, the maximum precision obtained is about 77%. When matching with criteria, the best result is obtained with criteria 1 and 2. On average a gain of 10% is obtained, with the best precision value being about 86%. Using a combination of all the criteria gives a slightly poorer performance, because the third criterion is too severe for the quality of the models obtained. Criteria 1 and 2 enhance the retrieval performance significantly. Hence, our model benefit from the use of additional information about the parts and their spatial relationships. Of course, this effect was expected, but it was nevertheless interesting to validate it experimentally.

Even more interesting is the study of the “precision vs recall” for the six objects individually. The results for this study are shown in Fig. 9. Our method performs better for simple objects like a watering can and a compass. It performs well also for the lamp. However, our method has difficulties with objects like the airplane and the stool. Our method performs poorly for the stool because a constant and good model cannot be obtained. This is because, our segmentation algorithm is not aware of the see-through holes in the stool. The inside area of the stool is not segmented from the background because only the external outline of the object is extracted. For the airplane,

the performances are poor because PLASTIQUE fails in extracting all the wings. Sometimes, because of scale, the wings in the image are thin and cause part segmentation difficulties. Hence, PLASTIQUE has an unwanted sensitivity to scale. Furthermore, the external outline often does not enclose all the wings because edge detection separates the wings from the fuselage. By solving these issues, the performance of PLASTIQUE could be significantly improved.

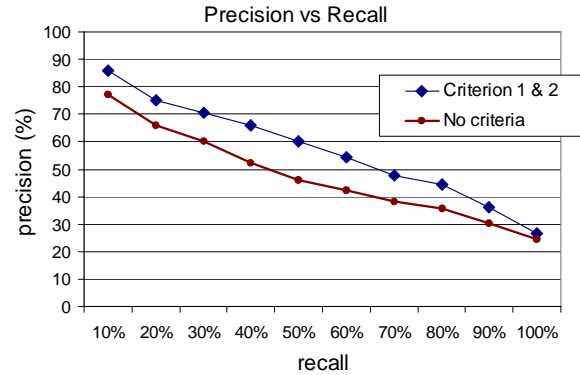


Fig. 8. Precision vs recall average for all queries when matching is conducted with no criteria and with a combination of criterion 1 and 2.

Globally, when observing Fig. 8 and Fig. 9, the results show that part-based models should not be neglected in image retrieval applications. First of all, even though not perfect, the current implemented algorithms can be applied for retrieving some types of objects, and this at the object level. Next, our method can be further improved by enhancing some modules, for example outline extraction. The outline extraction algorithm used is very simple. It could be easily improved. This way, object with holes or objects like airplanes could be segmented with more success.

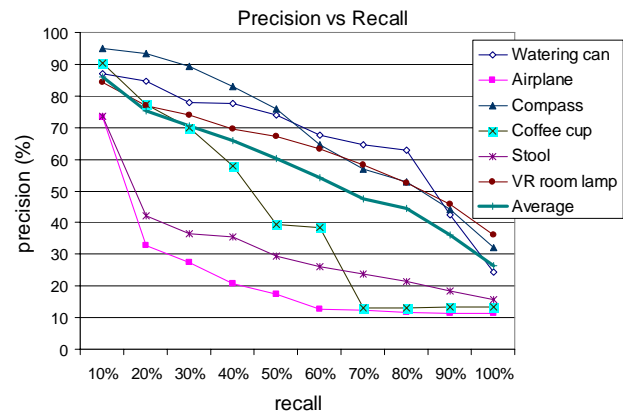


Fig. 9. Precision vs recall for the objects individually.

Part-based object models could be used in a specialized module in an image retrieval software, or in combination with other modeling approaches. For example, the image retrieval software of [5] has a module specialized for textures and another for shapes. Why not, a module for part-based objects? Also, used in combination, part-based modeling of objects could improve the retrieval of some type of images where other image models fail.

Another advantage of our part-based object model is that it allows formulating queries by drawing the outline of an object. Fig. 10 shows the results of a query formulated in this way using PLASTIQUE. The sketch is drawn by adding straight line segments and circular arcs.

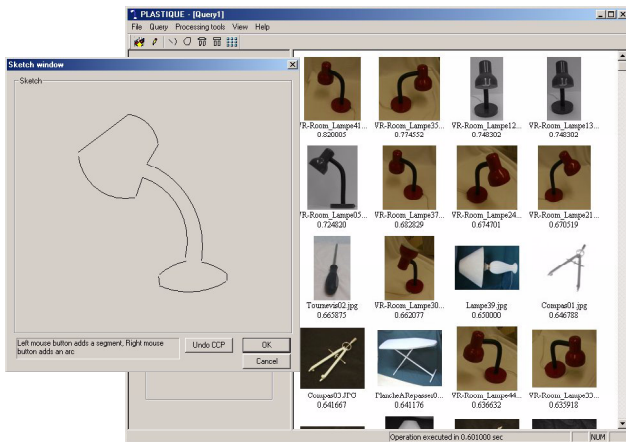


Fig. 10. Best 16 similar images obtained from a sketched query.

Conclusion

This paper has presented a 3D part-based method to model and match images in databases. This method allows images to be matched where objects have deformations or are seen from different viewpoints. It builds models directly from a single image. This is an important advantage as a single copy of an object can be matched without the need of explicitly building a model for each object.

Experiments demonstrate the abilities of our method, as PLASTIQUE successfully matches deformed objects and objects seen in different viewpoints. Current results show that PLASTIQUE is not as successful at matching different classes of objects. However, these results show that the method of PLASTIQUE has very good potential.

Analysis of the precisions for different recall values shows that our method can be further improved by extracting the internal outline of object and by improving segmentation of thin parts and outline extraction. These two issues will be the subject of future efforts. Additional experiments using more objects are currently underway.

Acknowledgements

This work is supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR).

References

- [1] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R., Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2000) 22(12), pp. 1349-1380.
- [2] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner Dom, J., Lee, D., Petkovic, D., Steele, D., Yanker, P., Query by image and video content: The QBIC system, *IEEE Computer*, (1995) 28(9), pp. 23--32.
- [3] Iqbal, Q., Aggarwal, J.K., Combining Structure, Color and Texture for Image Retrieval: A Performance Evaluation, *ICPR 2002*, Québec, Canada.
- [4] Hirata, K., Katsutani, E., Hara, Y., On Image Segmentation for Object-based Image Retrieval, *ICPR 2002*, Québec, Canada.
- [5] Pentland, A., Picard, R.W., Scarloff, S., Photobook: Content-Based Manipulation of Images Databases, *International Journal of Computer Vision*, (1996) 18(3), pp. 233-254.
- [6] Bilodeau, G.A., Bergevin, R., Generic Modeling of 3D Objects from Single 2D Images, in *International Conference on Pattern Recognition*, (2000), Barcelona, Spain. pp. 770-773
- [7] Bilodeau, G.A., Bergevin, R., Part segmentation of objects in real images, *Pattern Recognition*, vol. 35, November 2002, pp. 387-400.
- [8] Bilodeau, G.A., Bergevin, R., Modeling of 2D Parts Applied to Database Query, in *Vision interface 2001*, (2001), Ottawa, Canada. pp. 228-235.
- [9] Bergevin, R., Mokhtari, M., Multiscale Contour Segmentation and Approximation: An Algorithm Based on the Geometry of Regular Inscribed Polygons, *Computer Vision and Image Understanding*, (1998) (71(1)), pp. 55-75.
- [10] Biederman, I., Recognition-by-components: A theory of human image understanding. *Psychological Review*, (1987) 94, pp. 115-147.
- [11] Lowe, D.G., *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, 1985.
- [12] Bilodeau, G.A., Bergevin, R., Constructing and matching fuzzy graphs of volumetric primitives hypotheses, in *Vision Interface 2003*, (2003), Halifax, Canada.