Content-Based Image Retrieval Using Shape and Depth from an Engineering Database

Amit Jain, Ramanathan Muthuganapathy, and Karthik Ramani

School of Mechanical Engineering,
Purdue University,
West Lafayette, IN 47907, USA
{`amitj,rmuthuga,ramani`}@purdue.edu

Abstract. Content based image retrieval (CBIR), a technique which uses visual contents to search images from the large scale image databases, is an active area of research for the past decade. It is increasingly evident that an image retrieval system has to be domain specific. In this paper, we present an algorithm for retrieving images with respect to a database consisting of engineering/computer-aided design (CAD) models. The algorithm uses the shape information in an image along with its 3D information. A linear approximation procedure that can capture the depth information using the idea of shape from shading has been used. Retrieval of objects is then done using a similarity measure that combines shape and the depth information. Plotted precision/recall curves show that this method is very effective for an engineering database.

1 Introduction

Content-based image retrieval (CBIR), a technique which uses visual contents to search images from large scale image databases has been an active research area for the last decade. Advances in the internet and digital imaging have resulted in an exponential increase in the volume of digital images. The need to find a desired image from a collection of databases has wide applications, such as, in crime prevention by automatic face detection, finger print, medical diagnosis, to name a few.

Early techniques of image retrieval were based on the manual textual annotation of images, a cumbersome and also often a subjective task. Texts alone are not sufficient because of the fact that interpretation of what we see is hard to characterize by them. Hence, contents in an image, color, shape, and texture, started gaining prominence.

Initially, image retrievals used the content from an image individually. For example, Huang and Jean [1] used a 2D C^+ -strings and Huang et al. [2] used the color information for indexing and its applications. Approaches using a combination of contents then started gaining prominence. Combining shape and color using various strategies such as weighting [3], histogram-based [4], kernel-based [5], or invariance-based [6] has been one of the premier combination strategies.

Shape and texture using elastic energy-based approach to measure image similarity has been presented in [7]. Smith and Chang [8] presented an automated extraction of color and texture information using binary set representations. Li et al. [9] used a color histogram along with the texture and spatial information. Image retrieval by segmenting them had been the focus of few research papers such as [10] and [11]. A detailed overview on the various literatures that are available on CBIR can be found in [12] and [13]. A discussion on various similarity measurement techniques can be found in [14].

Even though research on image retrieval has grown exponentially, particularly in the last few years, it appears that less than 20% were concerned with applications or real-world systems. Though various combinations of contents and their possible descriptions have been tried, it is increasingly evident that a system cannot cater to the needs of a general database. Hence, it is more relevant to build image retrieval systems that are specialized to domains. Also, the selection of appropriate features for CBIR and annotation systems remain largely ad-hoc.

In this paper, retrieving images from an engineering database has been presented. As the engineering objects are geometrically well-defined as opposed to natural objects and also they rarely contain texture information, the appropriate features shape (or contour) to capture its two dimensional content along with its 3D embedding information, its depth profile at each pixel on the contour, has been used. Shape from an image is quite a powerful representation as it characterizes the geometry of the object. However, it is normally a planar profile, and is insufficient by itself to recognize objects that are typically 3D in nature. To take into account the third dimension, other parameters such as color and/or texture have been used. However, in our paper, we propose an approach that combines shape with the depth-map of the shape. The basic idea of our paper is illustrated in Fig. 1. Depth map, obtained from depth from focus approach using multiple images, has been used in [15] for indexing and retrieval by segmenting them. However, using depth information alone is not quite sufficient for well-defined geometric objects.

Fig. 1. Flow chart indicating the basic idea used in this paper

The rest of the paper is organized as follows. Section 2 describes the method to obtain the shape information, given an image. The method used for obtaining the 3D embedding information, i.e., the depth is described in Section 3. A representation involving both shape and depth along with its similarity measurements for retrieval is described in Section 4. Retrieval results are presented and discussed in Section 5. Finally, Section 6 concludes the paper.

2 Obtaining Shape

Engineering objects are geometrically well-defined as most of the them are obtained from a boolean combination of primitives. Hence, it is imperative to get the geometry information. As the input is an image, its 2D information can be obtained by applying a contour detection algorithm. This geometry information can be termed as the shape information for the particular image.

Steps to obtain the contour of an image are shown in Fig. 2. Contour can be obtained by separating the object information from its background details. This is done by converting the given image into a gray scale image (Fig. 2(a)). It is then binarized (Fig. 2(b)). As the contour detection algorithms are susceptible to small changes, converting to a binary image reduces its susceptibility. A simple threshold is applied to convert a gray scale image into the binary image. This conversion can induce noise along the shape boundary. Denoising using blurring techniques is then applied to remove them. It also eliminates isolated pixels, and small regions. Applying the contour tracing algorithm generates the boundary shape (contours) of the object (Fig. 2(c)). A polynomial is then fit to simplify the contours that generates the contour image.

Fig. 2. Processing an input image (a) Grayscale image (b) Binarized image (c) Contour extraction

Shape signature, a one dimensional representation of the shape, is obtained by applying the 8-point connectivity technique on the 2D closed contour. As engineering/CAD objects have well defined centroid (x_c, y_c), and also retrieval has shown to be better with central distance [16], we use it as our shape

representation. The feature vector representing the central distance between a point on the contour (x, y) and the centroid (x_c, y_c) is given by

$$V_c = (x - x_c, y - y_c, 0) \quad (1)$$

where $x_c = \frac{1}{N} \sum_{i=0}^{N-1} x_i$, $y_c = \frac{1}{N} \sum_{i=0}^{N-1} y_i$ and N is the total number of pixels.

3 Computing Depth Map

Once the shape or the contour is obtained (as described in Section 2), its 3D information is then computed. Recovering the 3D information can be done in terms of depth Z , the surface normal (n_x, n_y, n_z) , or surface gradient (p, q) . One approach is to use several images taken under different lighting conditions, as in Photometric stereo, and identify the depth by the change in illumination.

However, in this paper, we use only a single image and not a set of images. Hence, principles of shape from shading has been used to obtain the 3D embedding information. Lambertian model, where it is assumed that equal amount of light is reflected in every direction, is a reasonable approximation for engineering objects. In this model, the reflectance map is simplified to be independent of viewers direction. The important parameters in Lambertian reflectance are albedo, which is assumed to be constant and illuminant direction, which can be computed, in general.

To identify the depth-map (sometimes called as simply depth) of an image, we use the approach proposed by [17], where it is assumed that the lower order components in the reflectance map dominate. The linearity of the reflectance map in the depth Z has been used instead of in p and q . Discrete approximations for p and q are employed and linearize the reflectance in $Z(x, y)$. The following are used from [17] and has been presented here for completeness.

The reflectance function for the Lambertian surface is as follows:

$$E(x, y) = R(p, q) = \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2} \sqrt{1 + p_s^2 + q_s^2}} \quad (2)$$

where $E(x, y)$ is the gray level at pixel (x, y) , $p = \frac{\partial Z}{\partial x}$, $q = \frac{\partial Z}{\partial y}$, $p_s = \frac{\cos \tau \sin \sigma}{\cos \sigma}$, $q_s = \frac{\sin \tau \sin \sigma}{\cos \sigma}$, τ is the tilt of the illuminant and σ is the slant of the illuminant. Discrete approximation of p and q are given by the following:

$$p = \frac{\partial Z}{\partial x} = Z(x, y) - Z(x - 1, y), \quad q = \frac{\partial Z}{\partial y} = Z(x, y) - Z(x, y - 1) \quad (3)$$

The reflectance equation can be then rewritten as

$$\begin{aligned} 0 &= f(E(x, y), Z(x, y), Z(x - 1, y), Z(x, y - 1)) \\ &= E(x, y) - R(Z(x, y) - Z(x - 1, y), Z(x, y) - Z(x, y - 1)) \end{aligned} \quad (4)$$

For a fixed point (x, y) and a given image E , linear approximation (Taylor series expansion up through the first order terms) of the function f about a given depth

map Z^{n-1} and solving using iterative Jacobi method results in the following reduced form:

$$\begin{aligned} 0 &= f(Z(x, y)) \\ &= f(Z^{n-1}(x, y)) + (Z(x, y) - Z^{n-1}(x, y)) \frac{df(Z^{n-1}(x, y))}{dZ(x, y)} \end{aligned} \quad (5)$$

For $Z(x, y) = Z^n(x, y)$, the depth map at n -th iteration can be solved using the following:

$$Z^n(x, y) = Z^{n-1}(x, y) \frac{df(Z^{n-1}(x, y))}{dZ(x, y)} \quad (6)$$

where $\frac{df(Z^{n-1}(x, y))}{dZ(x, y)} = -1 * \left(\frac{p_s + q_s}{\sqrt{1+p^2+q^2}\sqrt{1+p_s^2+q_s^2}} - \frac{(p+q)(pp_s+qq_s+1)}{(\sqrt{1+p^2+q^2})^3\sqrt{1+p_s^2+q_s^2}} \right)$

Figs. 3(b) and 3(d) show the depth maps of the respective images in Fig. 3(a) and Fig. 3(c). It is to be noted that only the depth values at the contour are used in this paper even though they are calculated for the interior region as well. In general, Lambertian model assumption by itself is probably not sufficient. A more generalized model [18] that includes diffuse and specular properties can be used for better approximation of the depth map.

Fig. 3. Image and the depth maps

The depth map is then represented in a way similar to the shape (Equation (1)). The feature vector representing depth is given by

$$V_d = (0, 0, Z - Z_c) \quad (7)$$

where Z is the depth obtained from Equation (6) of the contour, and Z_c denotes the third dimension of the centroid.

4 Representation, Indexing and Retrieval

In this section, the introduced shape-depth representation is described, followed by Indexing using Fourier Descriptors and then a similarity measurement to describe the retrieval.

4.1 Shape-Depth Representation

As it is evident, shape alone is not sufficient to get a good retrieval. Typically, color has been combined with shape to obtain better retrieval results. As we are dealing with well-defined geometric objects, a novel strategy based on a 3D embedding has been adopted. Shape, in this paper, is combined with the corresponding estimated depth profile.

Shape-Depth can be defined as $I : R^2 \rightarrow R^3$. At each point on the contour, a vector is defined as follows:

$$V = (x - x_c, y - y_c, Z - Z_c) \quad (8)$$

Note that the vector Equation (8) is of dimension three, which is quite low and hence it enhances the speed of retrieving. This can be decomposed into V_c (Equation (1)) representing the shape/contour and V_d (Equation (7)) representing depth.

A weighted combination of the magnitude of the vectors V_c and V_d is used for retrieving images. Shape-Depth representation is defined as follows:

$$SD = \frac{w_c * \|V_c\| + w_d * \|V_d\|}{w_c + w_d} \quad (9)$$

where w_c and w_d are the weights assigned to the shape-based similarity and the depth-based similarity, respectively and $w_c + w_d = 1$, $w_c > 0$ and $w_d > 0$. It can be observed that $\|V_c\|$ captures the central distance measure in the 2D domain and $\|V_d\|$ is a similar measure on the third dimension, the depth. It should be noted that central distance captures both local and global features of the representation. It is safe to say that the shape-depth representation is between the contour-based and region-based representations and hence it could prove to be a very useful one for retrieving objects/images.

4.2 Fourier Transform of Shape-Depth and Indexing

A primary requirement of any representation for retrieval is that it is invariant to transformations such as translation, scaling and rotation. Fourier transform is widely used for achieving the invariance. For any 1-D signature function, its discrete Fourier transform is given by

$$a_n = \frac{1}{N} \sum_{i=0}^{N-1} (SD) \exp(-j2\pi it/N) \quad (10)$$

where, $n = 0, 1, \dots, N - 1$ and SD is given by Equation (9).

The coefficients a_n are usually called Fourier Descriptors (FD), denoted as FD_n . Since the shape and depth representations described in this paper are translation invariant, the corresponding FDs are also translation invariant. Rotation invariancy is achieved by using only the magnitude information and ignoring the phase information. Scale normalization is achieved by dividing the

magnitude values of the FDs with FD_1 . The invariant feature vector used to index SD is then given by

$$f = \left[\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|}, \dots, \frac{|FD_{N-1}|}{|FD_1|} \right] \quad (11)$$

4.3 Similarity Measurement

Retrieval result is not a single image but a list of images ranked by their similarities with the query image since CBIR is not based on exact matching. For a model shape indexed by FD feature $f_m = [f_m^1, f_m^2, \dots, f_m^N]$ and a database indexed by FD feature $f_d = [f_d^1, f_d^2, \dots, f_d^N]$, the Euclidean distance between two feature vectors can then be used as the similarity measurement:

$$d = \sqrt{\sum_{i=0}^{N-1} |f_m^i - f_d^i|^2} \quad (12)$$

where N is the total number of sampled points on the shape contour.

5 Experimental Results

For testing the proposed approach, the engineering database [19] containing 1290 images (a fairly good amount for testing), has been used. It has multiple copies of an image and also it has same images in arbitrary position and rotation. The query image is also one of the images in the databases. Our framework for CBIR is built in Visual C++. Test results for some objects are shown in Figs. 4(a) to 4(e), where only first fifteen retrieved images are shown for the query image on the right.

The various parameters that can influence the retrieval results in our approach are the following; number of sampling points to compute the FDs, weights w_c and w_d in Equation (9), and the factors such as light direction and number of iterations when computing the depth. There is always a tradeoff when the number of sampling points is chosen. Too large a value will give a very good result at the cost of computation. On the other hand, using lesser number of coefficients is computationally inexpensive but may not give an accurate information. Based on the experimentation conducted on the contour plots of the database, the number of coefficients was chosen to be 100. For this sampling, the values for the weights that yield better results were identified to be $w_c = 0.70$ and $w_d = 0.30$. The depth computation uses 2 iterations initialized to zero depth with light source direction as $(0, 0, 1)$. Fig. 4 shows the results for test images with the above mentioned parameters and the plotted precision-recall (Fig. 5) shows combined shape-depth representation yields better retrievals than using shape alone.

In all the test results, it is to be noted that the query image is also retrieved, which indicates that the shape-depth representation is robust. They also show that genus > 1 objects are retrieved when the query is of zero genus. This is

(a)

(b)

(c)

(d)

(e)

Fig. 4. Retrieval Results for some Engineering objects

Fig. 5. Precision-Recall for Shape and Shape-Depth representations

because of the following reasons: interior contour information is not used in this experiment and depth values only at the contour has been used. We believe that the retrieval results will be improved when the inside region of the contour along with the depth at the interior is used for shape-depth representation. We also did not carry out the experiment with various light source directions that could affect the depth-map.

The main advantage in using the depth content of the image is that we can represent objects close to how it is in its three dimensional space. As we are using only a single image to compute the depth map, it will be close to its real depth only if the image is in its most informative position. As a consequence, the current approach can produce very good results for objects having symmetry, such as 2.5D objects and also when the depth map is computed from the most informative position for general objects as is the case in most engineering images. However, in our approach, as we are not only dependent on the depth but also its shape information, we can also retrieve objects that are in different orientation as can be seen in Fig. 4(d), though we have not analyzed the bounds on the orientation. In future, the weights w_c and w_d can be identified dynamically based on the changes in the depth map and the shape-depth correspondence. A better representation for the obtained depth information is also being explored.

6 Conclusions

The main contribution of this paper is the idea of combining shape (contour) obtained from the contour tracing along with the 3D embedding, the depth information at each point on the contour. Similarity metrics are proposed to combine the shape and depth. It is shown that this approach is effective for retrieving engineering objects. It would be interesting to investigate whether the proposed shape representation is useful in other application domains, such as protein search in molecular biology.

References

1. Huang, P., Jean, Y.: Using 2d c+-strings as spatial knowledge representation for image database systems 27, 1249–1257 (1994)
2. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Spatial color indexing and applications. *Int. J. Comput. Vision* 35, 245–268 (1999)
3. Jain, A., Vailaya, A.: Image retrieval using color and shape. *Pattern Recognition* 29, 1233–1244 (1996)
4. Saykol, E., Gudukbay, U., Ulusoy, O.: A histogram-based approach for object-based query-by-shape-and-color in multimedia databases. Technical Report BU-CE-0201, Bilkent University, Computer Engineering Dept (2002)
5. Caputo, B., Dorko, G.: How to combine color and shape information for 3d object recognition: kernels do the trick (2002)
6. Diplaros, A., Gevers, T., Patras, I.: Combining color and shape information for illumination-viewpoint invariant object recognition 15, 1–11 (2006)
7. Pala, S.: Image retrieval by shape and texture. *PATREC: Pattern Recognition*. Pergamon Press 32 (1999)
8. Smith, J.R., Chang, S.F.: Automated image retrieval using color and texture. Technical Report 414-95-20, Columbia University, Department of Electrical Engineering and Center for Telecommunications Research (1995)
9. Li, X., Chen, S.C., Shyu, M.L., Furht, B.: Image retrieval by color, texture, and spatial information. In: *Proceedings of the the 8th International Conference on Distributed Multimedia Systems (DMS 2002)*, San Francisco Bay, CA, USA, pp. 152–159 (2002)
10. Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J.: Blobworld: A system for region-based image indexing and retrieval. In: *Third International Conference on Visual Information Systems*, Springer, Heidelberg (1999)
11. Shao, L., Brady, M.: Invariant salient regions based image retrieval under viewpoint and illumination variations. *J. Vis. Comun. Image Represent.* 17, 1256–1272 (2006)
12. Veltkamp, R., Tanase, M.: Content-based image retrieval systems: A survey. Technical Report UU-CS-2000-34, Utrecht University, Department of Computer Science (2000)
13. Datta, R., Li, J., Wang, J.Z.: Content-based image retrieval: approaches and trends of the new age. In: *MIR 2005: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*, pp. 253–262. ACM Press, New York (2005)
14. Chang, H., Yeung, D.Y.: Kernel-based distance metric learning for content-based image retrieval. *Image Vision Comput.* 25, 695–703 (2007)
15. Czúni, L., Csordás, D.: Depth-based indexing and retrieval of photographic images. In: García, N., Salgado, L., Martínez, J.M. (eds.) *VLBV 2003*. LNCS, vol. 2849, pp. 76–83. Springer, Heidelberg (2003)
16. Zhang, D.S., Lu, G.: A comparative study on shape retrieval using fourier descriptors with different shape signatures. In: *Proc. of International Conference on Intelligent Multimedia and Distance Education (ICIMADE 2001)*, Fargo, ND, USA, pp. 1–9 (2001)
17. Tsai, P., Shah, M.: Shape from shading using linear-approximation. *IVC* 12, 487–498 (1994)
18. Lee, K.M., Kuo, C.C.J.: Shape from shading with a generalized reflectance map model. *Comput. Vis. Image Underst.* 67, 143–160 (1997)
19. Jayanti, S., Kalyanaraman, Y., Iyer, N., Ramani, K.: Developing an engineering shape benchmark for cad models. *Computer-Aided Design* 38, 939–953 (2006)