23-10-2004

Image database retrieval using sketched queries

A. Chalechale
University of Wollongong

G. Naghdy
University of Wollongong, golshah@uow.edu.au

Prashan Premaratne
University of Wollongong, prashan@uow.edu.au

Recommended Citation

Chalechale, A.; Naghdy, G.; and Premaratne, Prashan: Image database retrieval using sketched queries 2004.
https://ro.uow.edu.au/infopapers/51

Image database retrieval using sketched queries

Abstract

This paper presents a novel approach for sketch-based image retrieval based on low-level features. It enables the measuring of the similarity among full color multi-component images within a database (models) and simple black and white user sketched queries. It needs no cost intensive image segmentation. Strong edges of the model image and morphologically thinned version of the query image are used for image abstraction. Angular-radial decomposition of pixels in the abstract images is used to extract new compact and affine invariant features. Comparative results, employing an art database (ArT BANK), show significant improvement in average normalized modified retrieval rank (ANMRR) using the proposed features.

Disciplines

Physical Sciences and Mathematics

Publication Details

This article was originally published as: Chalechale, A, Naghdy, G & Pramaratne, P, Image database retrieval using sketched queries, International Conference on Image Processing(ICIP '04), 24-27 October 2004, 1, 433-436. Copyright IEEE 2004.

IMAGE DATABASE RETRIEVAL USING SKETCHED QUERIES

Abdolah Chalechale, Golshah Naghdy, and Prashan Premaratne

School of Electrical, Computer & Telecommunications Engineering
University of Wollongong, Australia
{ac82,golshah,prashan}@uow.edu.au

ABSTRACT

This paper presents a novel approach for sketch-based image retrieval based on low-level features. It enables measuring the similarity among full color multi-component images within a database (models) and simple black and white user sketched queries. It needs no cost intensive image segmentation. Strong edges of the model image and morphologically thinned version of the query image are used for image abstraction. Angular-radial decomposition of pixels in the abstract images is used to extract new compact and affine invariant features. Comparative results, employing an art database (ArT BANK), show significant improvement in Average Normalized Modified Retrieval Rank (ANMRR) using the proposed features.

1. INTRODUCTION

With the advent of Internet and digital imaging devices such as scanners and digital cameras there is widespread access to virtually unlimited multimedia information. *Image* is one of the most important part of such data. Image retrieval from multimedia databases is still an open problem [1, 2]. There are many popular image retrieval techniques currently being used with different technologies. Representative content-based systems are Blobworld [3], QBIC [4], FourEyes [5], MetaSEEk and VisualSEEk [6, 7].

In sketch-based image retrieval, where the query example is a rough and simple black and white hand-drawn sketch, color and texture lose their original abilities to serve as the content key. Visual shape descriptors are useful in sketch-based image retrieval only when the model and the query images contain one object in a plain background [8]. In multiple-object scenes, object layout is a powerful tool, but object extraction and segmentation costs together with rotation variance introduce major drawbacks.

A huge number of researchers have considered the problem of content-based image retrieval (CBIR) while very few have addressed sketch-based image retrieval (SBIR). The work of Hirata and Kato, Query by Visual Example (QVE) [9], focuses solely on this problem. IBM has adopted a modified version of the approach in its QBIC system. In this approach the query and the target images are resized to 64×64 pixels and measuring the similarity is performed using a block correlation scheme. The approach does not allow indexing and is computationally expensive. Although the method can tolerate small local rotations, it is not rotation invariant and does not allow for large global rotations.

The edge histogram descriptor (EHD) was proposed in the visual part of the MPEG-7 standard [10]. It originally consists of an 80-bin histogram. A given image is divided into 16 sub-images (4×4) first, local edge histograms are then computed for each sub-image based on the strength of five different edges (vertical, horizontal, 45° diagonal, 135° diagonal, and isotropic). S. Won *et al.* proposed the efficient use of the descriptor by extending the histogram to 150 bins [11]. The extended histograms are obtained by further grouping the image blocks into 13 clusters (4 vertical, 4 horizontal, and 5 square clusters). A 65-bin semi-global histogram and a 5-bin global histogram are added to make a 150-bin histogram. The EHD is basically intended for gray-image to gray-image matching but changing the internal parameter Th_{edge}, a threshold for edge extraction, could make it applicable for black and white queries in sketch-based retrieval.

The 2-D Fourier transform in polar coordinates is employed for shape description. Its supremacy over 1-D Fourier descriptors, curvature scale space descriptor (MPEG-7 contour-based shape descriptor) and Zernike moments (MPEG-7 region-based descriptor) has been shown in [12]. Polar Fourier descriptor (PFD) is extracted from frequency domain by exploiting two-dimensional Fourier transform on polar raster sampled image. It is able to capture image features in both radial and spiral directions [12].

This paper focuses on the problem of finding image features, invariant to scale, translation, and rotation, which can be used efficiently in sketch-based retrieval. We also eliminate any constraint regarding the number of objects and the existence of any background. In other words, the images may have several objects in an inhomogeneous background. Applications for this approach include image search engines, in particular, for art gallery search or web-based image search. Image abstraction and spatial decomposition are used for feature extraction. An art image experimental database, called ArT BANK, has been developed and employed for test in this study. Average Normalized Modified Retrieval Rank (ANMRR), which is defined in MPEG-7 standard [13], is used for performance evaluation.

The details of the proposed approach are discussed in the next section. Section 3 exhibits experimental results, evaluation, and discussion. Section 4 concludes the paper and poses some new directions.

0-7803-8554-3/04/$20.00

2. ABSTRACT IMAGE DECOMPOSITION AND TRANSFORM (AIDT)

The main objective of the proposed approach is to transform the image data into a new structure that supports measuring the similarity between a full color image and a black and white hand-drawn simple sketch. The proposed algorithm is scale, translation and rotation invariant.

The edge map of an image carries the solid structure of the image independent of the color attributes. Its applicability is well known in computer vision, pattern recognition and image retrieval. Edges are also proven to be a fundamental primitive of an image for the preservation of both the semantics and the perceived attributes. Furthermore, in sketch-based image retrieval, it is the most useful feature that can be employed for matching purposes [14].

According to the assumption that sketched queries are more similar to the edge maps, which contain only the perceptive and vigorous edges, we obtain two abstract images through the strong edges of the model image and using the thinned version of the query image. The proposed features are then extracted from the derived abstract images.

The full color model image is initially converted to a gray intensity image I, by eliminating the hue and saturation while retaining the luminance. The edges are then extracted using the Canny operator with $\sigma = 1$ and Gaussian mask of size $= 9$ using the following procedure for depicting the most perceived edges.

The values of high and low thresholds for the magnitude of the potential edge points are automatically computed in such a way that only the strong edges remain. This improves the general resemblance of the resulted edge map and the hand drawn query. In order to depict strong edges, let G be the Gaussian 1-D filter and g be the derivative of the Gaussian used in the Canny edge operator. Then

$$H(k) = \sum_{i} G(i)g(k+1-i)$$

is the 1-D convolution of the Gaussian and its derivative.

$$X(u,v) = [\sum_{j=1}^{V} I'(u,j)H(v-j)]'$$

and

$$Y(u,v) = \sum_{i=1}^{U} I(i,v)H(u-i)$$

for $u = 1,2,3,\ldots U$ and $v = 1,2,3,\ldots V$, are the vertical and horizontal edge maps respectively. U is the number of rows and V is the number of columns in the image I. The $'$ notation indicates matrix transpose. The magnitude of the edge points is then obtained as

$$\Gamma(u,v) = \sqrt{X(u,v)^2 + Y(u,v)^2}.$$

For efficient selection of the high and low thresholds, we then make a 64-bin cumulative histogram of the $\Gamma(u,v)$ values and find the minimum index ι in this cumulative histogram that is grater than $\alpha * U * V$, where α denotes the percentage of non-edge points in the image ($\alpha = 0.7$ is an adequate choice for many images). To retain strong edges of the image, $* \iota$ is selected as the high threshold

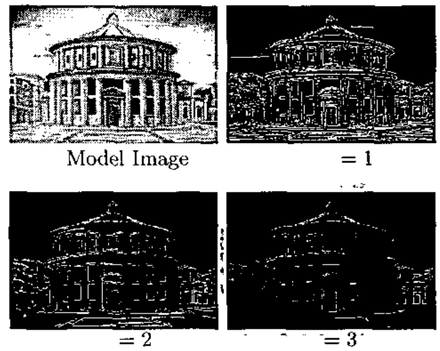

Fig. 1. The effect of parameter on the Canny edges at one of the art work images

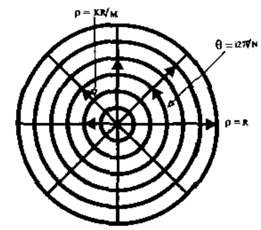

Fig. 2. Angular-radial partitioning to N angular and M radial sectors where $k = 0,1,2\ldots M-1$ and $i = 0,1,2\ldots N-1$

value and $0.4 \; * \iota$ is used for the low threshold value in the Canny edge operator. is a parameter that controls the degree of the strength of the edge points. Higher 's lead to lower number of edge points but more perceptive ones (see Fig. 1). Consequently, the gray image I is converted to edge image P using the Canny edge operator exploiting the above automatic extracted thresholds.

For query images, the procedure of black and white morphological thinning is applied to extract a thinned version of the sketched image. This image, namely Q, shows an outline of the query and contains the main structure of the user request. It contains spatial distribution of pixels similar to the strong edge map of the model image P.

The bounding box of P and Q images are then normalized to $J \times J$ pixels, using nearest neighbor interpolation. The proposed normalization of P and Q images ensures the scale and translation invariant properties. The resulting images are called *abstract image* Ω and used for the feature extraction scheme.

Based on the fact that any rotation of a given image, with respect to its center, moves a pixel at $(\rho, \;)$ to a new position at $(\rho, \; + \tau)$, we define a set of adjacent sectors in the abstract image Ω. In the following, we consider pixels $\Omega(\rho, \;)$ to be either equal to "1" for edge pixels or "0" for non-edge pixels.

The algorithm uses the surrounding circle of Ω for partitioning it to $M \times N$ sectors, where M is the number of radial partitions and N is the number of angular partitions. The angle between adjacent angular partitions is $\; = 2\pi/N$ and the radius of successive concentric circles is $\rho = R/M$

where R is the radius of the surrounding circle of the image (see Fig. 2).

The number of edge points in each sector of I is chosen to represent the sector feature. The scale invariant image feature is then $\{f(k,i)\}$ where

$$f(k,i)=\sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}}\ \sum_{\ =\frac{i2\pi}{N}}^{\frac{(i+1)2\pi}{N}}\Omega(\rho,\ ) \qquad (1)$$

for $k=0,1,2\ldots M-1$ and $i=0,1,2\ldots N-\ 1$. The features extracted above will be circularly shifted when the image Ω is rotated $\tau=l2\pi/N$ radians ($l=0,1,2\ldots$). To show this, let Ω denote the abstract image Ω after rotation by τ radians in counterclockwise direction:

$$\Omega\ (\rho,\ )=\Omega(\rho,\ \ -\tau). \qquad (2)$$

Then,

$$f\ (k,i)=\sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}}\ \sum_{\ =\frac{i2\pi}{N}}^{\frac{(i+1)2\pi}{N}}\Omega\ (\rho,\ ) \qquad (3)$$

are the image feature elements for Ω for the same k and i. We can express f as

$$\begin{aligned} f\ (k,i) &= \sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}}\sum_{\ =\frac{i2\pi}{N}}^{\frac{(i+1)2\pi}{N}}\Omega(\rho,\ \ -\tau) \\ &= \sum_{\rho=\frac{kR}{M}}^{\frac{(k+1)R}{M}}\sum_{\ =\frac{(i-l)2\pi}{N}}^{\frac{(i-l+1)2\pi}{N}}\Omega(\rho,\ ) \qquad (4) \\ &= f(k,i-l) \end{aligned}$$

where $i-l$ is a modulo M subtraction. It means that there is a circular shift (for individual k's) in the image features $\{f\ (k,i)\}$ compared to the image features $\{f(k,i)\}$ representing Ω and Ω respectively.

Using 1-D discrete Fourier transform of $f(k,i)$ and $f\ (k,i)$ for each k we obtain

$$\begin{aligned} F(k,u) &= \tfrac{1}{N}\textstyle\sum_{i=0}^{N\ 1} f(k,i)e^{\ j2\pi ui/N} \\ F\ (k,u) &= \tfrac{1}{N}\textstyle\sum_{i=0}^{N\ 1} f\ (k,i)e^{\ j2\pi ui/N} \\ &= \tfrac{1}{N}\textstyle\sum_{i=0}^{N\ 1} f(k,i-l)e^{\ j2\pi ui/N} \qquad (5) \\ &= \tfrac{1}{N}\textstyle\sum_{i=\ l}^{N\ 1\ l} f(k,i)e^{\ j2\pi u(i+l)/N} \\ &= e^{\ j2\pi ul/N}F(k,u). \end{aligned}$$

Because of the property $|F(k,u)|=|F\ (k,u)|$, the scale and rotation invariant image features are chosen as $\Upsilon=\{|F(k,u)|\}$ for $k=0,1,2\ldots M-1$ and $u=0,1,2\ldots N-1$.

Choosing a medium-size sector (e.g. $M=7$ and $N=10$) makes the invariant image feature Υ to be robust to other small variations as well (i.e. translation, erosion and occlusion). This is based on the fact that the number of edge pixels in such sectors varies slowly with such variations. Fig. 3 shows an example of Ω image superimposed with angular-radial partitions.

The similarity between images is measured by the ℓ_1 (Manhattan) distance between the two corresponding feature vectors. Experimental results (Section 3) confirm the robustness and efficacy of the extracted features.

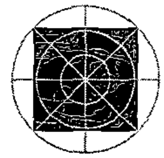

Fig. 3. An example of abstract image Ω and the Angular-radial partitions used for the feature extraction

3. EXPERIMENTAL RESULTS AND EVALUATION

This section presents experimental results using the new proposed algorithm in comparison with three other methods known from the literature. We made a collection of different model and query images called ArT BANK. Currently, it contains 3600 full color images of various sizes in the model part and 220 sketches in its query part. Images in the model part are mostly art works obtained from the World Art Kiosk at California State University. Each image was rotated 45°, 90°, 135°, 180°, 225°, 270° and 315° and put in the database. Images in the query part are black and white images, which are draft sketches drawn by different subjects similar to 55 arbitrary candidates from the model part and their rotated versions (90°, 180° and 270°). This is to evaluate rotation invariance property and to simulate different vertical and horizontal directions when posing a sketched query to the retrieval system. Fig. 4 shows an example of sketched image and the corresponding model image.

The ArT BANK was used as the test bed for the following four approaches: the proposed AIDT method (Section2), the QVE approach, as it used in the QBIC system [4], MPEG-7 edge histogram descriptor (EHD) [11], and the polar Fourier descriptor (PFD) approach [12]. All methods were tested using the images in the query part as input queries while regarding the images in the model part as database entries.

In the AIDT method (Section 2), we applied $M=7$, $N=10$, $\ =3$, and $J=129$ resulting in a 70-entry feature vector Υ. For the EHD method, *desired_num_of_blocks* was set to 1100 and Th_{edge} set to 11 (the default values) for the model images and Th_{edge} was set 0 for the queries since they are binary images. A 150-bin histogram was obtained in this approach employing local, semi-global and global histograms. We followed the algorithm given in [12] to obtain a 60-bin feature vector in PFD approach. The quantization stage in the EHD method was dropped to put all methods in a similar footing.

The ℓ_1 distance was used for measuring the similarity between image features of the MPEG-7 edge histogram descriptor (EHD) approach and of the proposed AIDT method. For the EHD method a weighting factor of 5 for global bins, as recommended in [11], was applied. Euclidian distance (ℓ_2) was exploited for measuring the similarity between PFD features [12] and global correlation factor was employed for measuring the similarity between images in the QVE method [4]. The generated vectors of AIDT, EHD, and PFD methods are used to represent the images. Average Normalized Modified Retrieval Rank (ANMRR),

Fig. 4. An example of sketched image (left) and corresponding model image (right)

which was developed during the MPEG-7 standardization activity [13], is used as retrieval performance measure. It exploits not only the recall and precision information but also the rank information among the retrieved images. Note that NMRR and its average (ANMRR) will always be in the range of [0, 1]. Based on the definition of ANMRR, the smaller the ANMRR, the better the retrieval performance.

Table 1 exhibits ANMRR for QVE, EHD, PFD, and AIDT methods. The retrieval performance of the proposed AIDT approach is the best (0.3501). PFD method's performance is better than EHD and QVE methods respectively (0.3935<0.4112<0.4918). It means the PFD approach is more applicable in sketch-based image retrieval than the other two approaches. QVE method exhibits lack of rotation invariance. This arises from the fact that the method looks only at local features and ignores global translation and rotation.

Since AIDT, EHD, and PFD methods generate a feature vector for each image, they support indexing. On the other hand, QVE approach ,which uses a correlation scheme for measuring the similarity between images, cannot be used to generate indices for the database.

Table 1. Retrieval performance of different methods

Method	QVE	EHD	PFD	AIDT
ANMRR	0.4918	0.4112	0.3935	0.3501

4. CONCLUDING REMARKS

The image analyzing approach presented in this paper (AIDT) enables measuring the similarity between a full color multi-component model image and a black and white sketched query. The images are arbitrary and may contain several complex objects in inhomogeneous backgrounds. The approach deals directly with the whole image and needs no cost intensive image segmentation and object extraction. Two abstract images are defined, based on the strong edges of the model image and the thinned outline of the query image respectively, and used for feature extraction. Angular-radial decomposition of the abstract images are used to extract features that are scale, translation, and rotation invariant. Experimental results, using the AIDT approach and the ArT BANK, as the test bed, show significant improvement in the ANMRR measure over three other well-known approaches within the literature. The decomposition scheme could be refined to improve retrieval performance. We intend to investigate sketch-based object recognition using the AIDT approach.

Acknowledgments. The authors would like to acknowledge the World Art Kiosk at California State University for providing paintings used in the ArT BANK. We also thank Dr. Nargess Yasini, Andrew Mani and Khadijeh Moosavian who kindly help us producing sketched queries. The first author is financially supported by the Ministry of Science, Research, and Technology of Iran.

5. REFERENCES

[1] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 22, no. 11, pp. 1349–1380, 2000.

[2] Y. Rui, T.S. Huang, and S.F. Chang, "Image retrieval: current techniques, promising directions, and open issues," *Journal of Visual Commun. and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.

[3] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 24, no. 8, pp. 1026–1038, 2002.

[4] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC project: querying images by content using color, texture, and shape," in *Proc. of Spie*, USA, 1993, vol. 1908, pp. 173–187.

[5] T. P. Minka and R. W. Picard, "Interactive learning with a "society of models"," *Patt. Recog.*, vol. 30, no. 4, pp. 565–581, Apr. 1997.

[6] M. Beigi, A. B Benitez, and S. Chang, "MetaSEEk: a content-based meta-search engine for images," in *Proc. of Spie*, USA, 1997, vol. 3312, pp. 118–128.

[7] J. R. Smith and S. Chang, "VisualSEEk: a fully automated content-based image query system," in *Proc. ACM Multimedia 96.*, NY, USA, 1996, pp. 87–98.

[8] H. H. S. Ip, A. K. Y. Cheng, W. Y. F. Wong, and J. Feng, "Affine-invariant sketch-based retrieval of images," in *Proc. IEEE Int. Conf. Comput. Graphics*, 2001, pp. 55–61.

[9] K. Hirata and T. Kato, "Query by visual example-content based image retrieval," in *Advances in Database Technology - EDBT '92*, Berlin, Germany, 1992, pp. 56–71.

[10] ISO/IEC JTC1/SC29/WG11/N4063, "MPEG-7 Visual part of eXperimentation Model Version 10.0," Singapore, March 2001.

[11] C. S. Won, D. K. Park, and S. Park, "Efficient use of Mpeg-7 edge histogram descriptor," *Etri Journal*, vol. 24, no. 1, pp. 23–30, Feb. 2002.

[12] D. Zhang and G. Lu, "Shape-based image retrieval using generic fourier descriptor," *Signal Processing: Image Commun.*, vol. 17, no. 10, pp. 825–848, 2002.

[13] ISO/IEC JTC1/SC29/WG11-MPEG2000/M5984, "Core experiments on MPEG-7 edge histogram descriptors," Geneva, May 2000.

[14] A. Chalechale and A. Mertins, "An abstract image representation based on edge pixel neighborhood inform. (EPNI)," *Lecture Notes in Computer Science*, vol. 2510, pp. 67–74, 2002.