SKETCH-BASED IMAGE RETRIEVAL USING ANGULAR PARTITIONING

Abdolah Chalechale, Golshah Naghdy

University of Wollongomg
Wollongong, Australia

Alfred Mertins

University of Oldenburg
Oldenburg, Germany

ABSTRACT

This paper presents a novel approach for sketch-based image retrieval based on low-level features. The approach enables measuring the similarity between a full color image and a simple black and white sketched query and needs no cost intensive image segmentation. The proposed method can cope with images containing several complex objects in an inhomogeneous background. Abstract images are obtained using strong edges of the model image and thinned outline of the sketched image. Angular-spatial distribution of pixels in the abstract images is then employed to extract new compact and effective features using Fourier transform. The extracted features are scale and rotation invariant and robust against translation. A collection of paintings and sketches (ART BANK) is used for testing the proposed method. The results are compared with three other well-known approaches within the literature.

Experimental results show significant improvement in the Recall ratio using the new features.

1. INTRODUCTION

The expeditious growth of digital imaging technology has resulted in the availability of a constantly increasing number of digital images. Image retrieval from multimedia databases is still an open problem. Traditional textual methods have been shown to be inefficient and insufficient for searching in visual data. Consequently, image and video content-based indexing and retrieval methods have attracted many new researchers in recent years [1, 2].

Existing content-based image retrieval systems include QBIC [3], PicSOM [4], MetaSEEk and VisualSEEk [5, 6]. MPEG-7 standard defines descriptors driven from three main image content features, i.e. color, texture and shape [7, 8] and VisualSEEk is using object layout as another content key [6].

User interaction is one of the most important aspects of any multimedia system. A simple user interface makes the system more attractive and applicable. In sketch-based image retrieval, when the query example is a rough and simple black and white draft, hand drawn by the user, color and texture lose their original ability to serve as content keys. Visual shape descriptors are useful in sketch-based image retrieval when the model and the query image contain only one object in a plain background [9]. In multiple-object scenes, object layout is a powerful tool, but object extraction and segmentation costs and rotation variance are the drawbacks.

We focus our discussion on the problem of finding image features, invariant to scale, translation and rotation, which can be used efficiently in sketch-based retrieval where the images have several objects in an inhomogeneous background. To our knowledge, the work of Hirata and Kato, Query by Visual Example (QVE) [10], is the only one that addressed this problem previously. IBM adopts a modified version of the approach in the QBIC system [3]. In this approach the query and the target images are resized to 64×64 pixels first and then 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 [11]. It consists of a 150-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). 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 in [12]. 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 [13].

Edge Pixel Neighborhood Information (EPNI) is utilizing neighborhood structure of the edge pixels to make an extended feature vector [14]. The vector is used efficiently for measuring the similarity between sketched queries and arbitrary model images. The semantic power of the method is examined in [15]. Although the method is scale and translation invariant it does not exhibit rotation invariance property.

In this paper we present a novel approach of feature extraction for matching purposes based on angular partitioning of an abstract image. Abstract images are obtained from the model image and from the query image by two different procedures. The angular-spatial distribution of pixels in the abstract image is then employed as the key concept for feature extraction using Fourier transform. The extracted features are scale and rotation invariant and robust against translation. The major contribution of the paper is in rotation invariance and translation robustness properties as, based on the existing literature, there is no rotation and translation invariant approach dealing with arbitrary images containing multiple objects while using sketched queries. The proposed algorithm and three other well-known approaches are implemented and examined using an art-image database (ART BANK). Experimental results show significant improvement in the Recall ratio using the new approach.

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

2. ANGULAR PARTITIONING OF ABSTRACT IMAGE (APAI)

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 edge map of an image carries the solid structure of the image independent of the color attribute. Edges are also proven to be a fundamental primitive of an image for the preservation of both semantics and perceived attributes [16]. Furthermore, in sketch-based image retrieval, edges form the most useful feature for matching purposes [14, 9, 17].

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 the thinned version of the query image. The proposed features are then extracted from the 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 semblance 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 let g be the derivative of the Gaussian used in the Canny edge operator. Then

$$H(k) = \sum_i G(i)g(k-i)$$

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

$$X(m,n) = [\sum_{j=1}^{N} I'(m,j)H(n-j)]'$$

$$Y(m,n) = \sum_{i=1}^{M} I(i,n)H(m-i)$$

for $m = 1,2,3,\dots M$ and $n = 1,2,3,\dots N$ are the vertical and horizontal central edge maps respectively, where M is the number of rows and N is the number of columns of the image I. The magnitude of the edge points is then obtained as

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

For efficient selection of the high and low thresholds, we make a 64-bin cumulative histogram of the $\Gamma(m,n)$ values and find the minimum index ι in this cumulative histogram that is grater than $\alpha * M * N$, 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 abstract image, $\beta * \iota$ is selected as the high threshold value and $0.4\beta * \iota$ is used for the low threshold value in the Canny edge operator. β is a parameter that controls the degree of the strength of the edge points. Higher β's lead to lower number of edge points but more perceptive ones (see Figure 1). Consequently, the gray image I is converted to an edge image P using the Canny edge operator exploiting the above automatic extracted thresholds.

For the query images, the procedure of black and white morphological thinning [18] is applied to extract a thinned version of

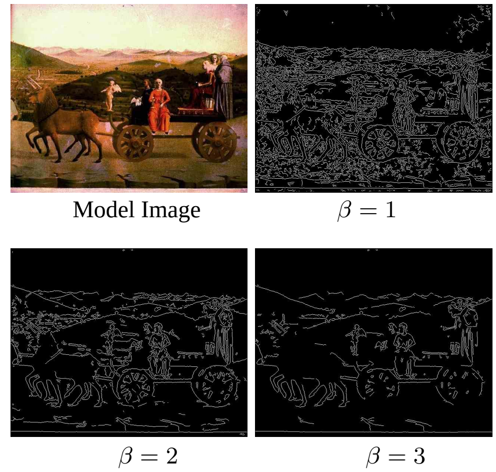

Figure 1: The effect of the β parameter on the edge maps.

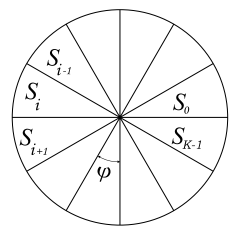

Figure 2: Angular partitioning splits the abstract image into K successive slices.

the sketched image. This image, namely Q, shows an outline of the query and contains the main structure of the user request. It contains the spatial distribution of pixels similar to the strong edge map of the model image P. The area in the bounding box of the comparable P and Q images are normalized to $J \times J$ pixels using nearest neighbor interpolation.The proposed normalization of the P and Q images ensures the scale invariance and translation robustness properties. The resulted image is called *abstract image* Ω.

We then define angular partitions (slices) in the surrounding circle of the abstract image Ω (see Figure 2). The angel between adjacent slices is $\varphi = 2\pi/K$ where K is the number of angular partitions in the abstract image. K can be adjusted to achieve hierarchical coarse to fine representation requirement. Any rotation of a given image, with respect to its center, moves a pixel at slice S_i to a new position at slice S_j where $j = (i+\lambda) \bmod K$, for $i, \lambda = 0,1,2,\dots K-1$. Figure 3 shows an example of abstract image Ω and its angular partitions.

The number of edge points in each slice of Ω is chosen to represent the slice feature. The scale and translation invariant image feature is then $\{f(i)\}$ where

$$f(i) = \sum_{\theta=\frac{i2\pi}{K}}^{\frac{(i+1)2\pi}{K}} \sum_{\rho=0}^{R} \Omega(\rho,\theta)$$

for $i = 0,1,2\dots K-1$. R is the radius of the surrounding circle of the abstract image. The feature extracted above will be circularly shifted when the image Ω is rotated $\tau = l2\pi/K$ radian ($l = 0,1,2\dots$). To show this, let Ω_τ denote the abstract image Ω after rotation by τ radians in counterclockwise direction:

$$\Omega_\tau(\rho,\theta) = \Omega(\rho,\theta-\tau).$$

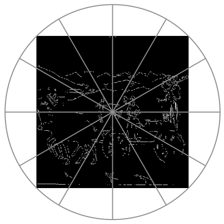

Figure 3: An example of abstract image Ω and the angular partitions (slices) used for the feature extraction.

Then,

$$f_\tau(i) = \sum_{\theta=\frac{i2\pi}{K}}^{\frac{(i+1)2\pi}{K}} \sum_{\rho=0}^{R} \Omega_\tau(\rho,\theta)$$

are the image feature elements of the Ω_τ for the same i's. We can express f_τ as

$$\begin{aligned} f_\tau(i) &= \textstyle\sum_{\theta=\frac{i2\pi}{K}}^{\frac{(i+1)2\pi}{K}} \sum_{\rho=0}^{R} \Omega(\rho,\theta-\tau) \\ &= \textstyle\sum_{\theta=\frac{(i-l)2\pi}{K}}^{\frac{(i-l+1)2\pi}{K}} \sum_{\rho=0}^{R} \Omega(\rho,\theta) \\ &= f(i-l) \end{aligned}$$

where $i-l$ is a modulo M subtraction. $f_\tau(i) = f(i-l)$ indicates that there is a circular shift in the image feature $\{f_\tau(i)\}$ regarding to the image feature $\{f(i)\}$ which representing Ω_τ and Ω respectively.

Using 1-D discrete Fourier transform of $f(i)$ and $f_\tau(i)$ we obtain

$$\begin{aligned} F(u) &= \tfrac{1}{K}\textstyle\sum_{i=0}^{K-1} f(i)e^{-j2\pi ui/K} \\ F_\tau(u) &= \tfrac{1}{K}\textstyle\sum_{i=0}^{K-1} f_\tau(i)e^{-j2\pi ui/K} \\ &= \tfrac{1}{K}\textstyle\sum_{i=0}^{K-1} f(i-l)e^{-j2\pi ui/K} \\ &= \tfrac{1}{K}\textstyle\sum_{i=-l}^{K-1-l} f(i)e^{-j2\pi u(i+l)/K} \\ &= e^{-j2\pi ul/K}F(u). \end{aligned}$$

Based on the property $\|F(u)\| = \|F_\tau(u)\|$, the scale and rotation invariant image features are chosen as $\{\|F(u)\|\}$ for $u = 0, 1, 2 \ldots K-1$. The features are also robust against translation because of the proposed normalization of the abstract images.

The similarity between the model and the sketched images is measured by the ℓ_1 (Manhattan) distance between the two feature vectors. Experimental results (Section 3) confirm the robustness and efficiency of the method.

3. EXPERIMENTAL RESULTS AND EVALUATION

This section presents experimental results using the new proposed approach 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 365 full color images of various sizes in the model part and 180 sketches in its query part. Images in the model part are mostly art works gained from the World Art Kiosk at California State University. Images in the query part are black & white images which are draft sketches similar to 45 arbitrary candidates from the model part and their rotated versions (90°, 180° and 270°).

The ART BANK was used as the test bed for the following four approaches. The proposed method (APAI), the QVE approach, as it used in the QBIC system [3], MPEG-7 Edge Histogram Descriptor (EHD) [**?**], and polar Fourier descriptor (PFD) approach [13]. All the methods were tested using the same database.

In the APAI method (Section 2), we applied $\beta = 3$, $J = 129$ and $\varphi = 5^\circ$, resulting in a 72-entry feature vector f. In 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 as 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 [13] to obtain a 60-bin feature vector in PFD approach. The quantization stage in the EHD method was ignored to put all methods in the same situation.

The queries were depicted from the following sets (each contains 45 images): the original sketches (Q_0), their $90^\circ(Q_{90})$, 180° (Q_{180}) and the $270^\circ(Q_{270})$ rotated versions. This is to simulate different vertical and horizontal directions when posing a sketched query to the retrieval system.

Recall ratio R_n [10] was used for the evaluation of retrieval performance. It shows the ratio to retrieve the original full color model image in the best n-candidates. That is

$$R_n = \frac{\text{queries retrieved the target image in the first } n \text{ retrievals}}{\text{total number of queries}} * 100$$

The R_n was obtained for each approach using the four different query sets. Figure 4 exhibits the retrieval performance supremacy of the proposed APAI method over the other approaches. For the queries with the same directions as the model images (Q_0 set), the retrieval performance of QVE, PFD and EHD methods decline respectively compared to the APAI method for almost all n's. The Recall ratio R_n of the PFD method is higher than QVE and EHD in all rotated cases. QVE and EHD methods show missing rotation invariance while PFD method is more robust with respect to rotation and the APAI exhibits the best rotation invariant property. The maximum R_n of the PFD method is 95% for $n = 100$ in set Q_{180} but R_n of the APAI reaches to 100% for $n \geq 50$ in all sets.

4. CONCLUSIONS

The approach presented in this paper (APAI) enables measuring the similarity between a full color model image and a simple black & 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. Abstract images were defined based on the strong edges of the model image and the morphological thinned outline of the query image. Angular partitioning of the abstract image, using Fourier transform, is exploited to extract features that are scale and rotation invariant and robust against translation. Experimental results, using APAI approach and the ART BANK as the test bed, show significant improvement in the Recall ratio over three other approaches known from the literature.

The partitioning scheme could be refined to improve retrieval performance. We also intend to investigate sub-image search by dividing the image into several regions and applying the approach to each region successively.

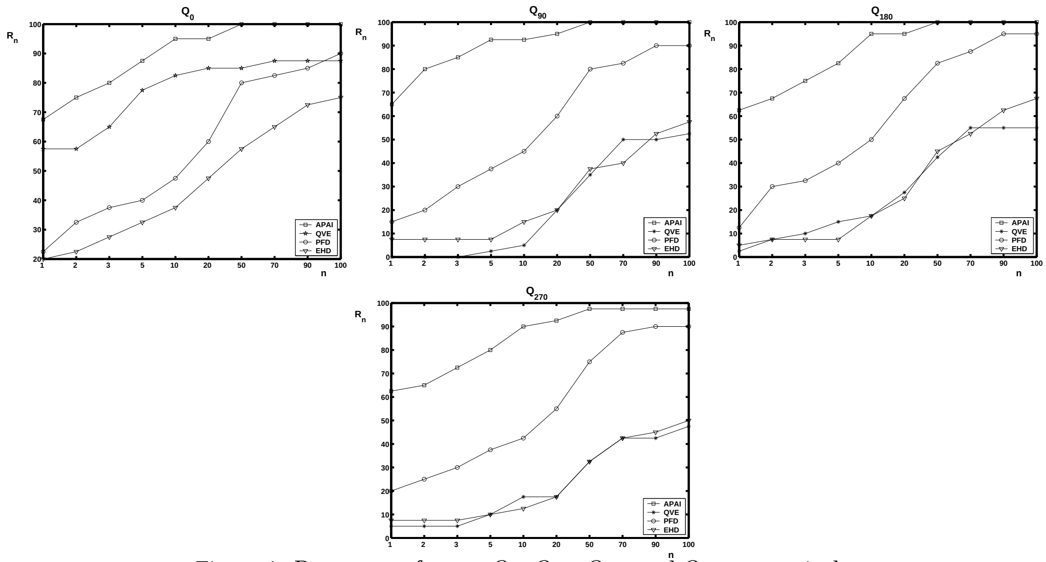

Figure 4: R_n verses n for sets Q_0, Q_{90}, Q_{180} and Q_{270} respectively.

Acknowledgments. The authors would like to acknowledge the World Art Kiosk at California State University for providing the art works used in this study. We also thank Dr. Nargess Yasini, Andrew Mani and Khadijeh Moosavian who kindly help us producing sketched queries.

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. on 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] 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.

[4] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "PicSOM-content-based image retrieval with self-organizing maps," *Patt Recognition Letters*, vol. 21, no. 13-14, pp. 1199–1207, Dec. 2000.

[5] 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.

[6] 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.

[7] B. S. Manjunath, J.-R. Ohm, and V. V. Vasudevan, "Color and texture descriptors," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 11, no. 6, pp. 703–715, June 2001.

[8] M. Bober, "Mpeg-7 visual shape descriptors," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 11, no. 6, pp. 716–719, June 2001.

[9] 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. on Comput. Graphics*, 2001, pp. 55–61.

[10] 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.

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

[12] D. Zhang and L. Guojun, "Generic fourier descriptor for shape-based image retrieval," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, 2002, vol. 1, pp. 425–428.

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

[14] A. Chalechale and A. Mertins, "An abstract image representation based on edge pixel neighborhood inform. (EPNI)," in *Lecture Notes in Comput. Science, EurAsian-ICT 2002:Inform. and Commun. technology*, 2002, vol. 2510, pp. 67–74.

[15] A. Chalechale and A. Mertins, "Semantic evaluation and efficiency comparison of the edge pixel neighboring histogram in image retrieval," in *WITSP'02 first workshop on the internet, telecommunications and signal processing*, Australia, 2002, pp. 179–184.

[16] L. Atzori and F.G.B. De Natale, "Error concealment in video transmission over packet networks by a sketch-based approach," *Signal Processing: Image Commun.*, vol. 15, no. 1-2, pp. 57–76, 1999.

[17] S. Matusiak, M. Daoudi, T. Blu, and O. Avaro, "Sketch-based images database retrieval," in *Advances in Multimedia Inform. Syst. 4th Int. Workshop, MIS'98. Proc.*, 1998.

[18] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, 1992.