OBJECT RETRIEVAL BASED ON USER-DRAWN SKETCHES

Sang Min Yoon and Arjan Kuijper
GRIS TU Darmstadt, Fraunhofer IGD & TU Darmstadt, Fraunhoferstrasse 5, Darmstadt, 64283, Germany

Keywords: Diffusion tensor fields, Sketch query based image retrieval.

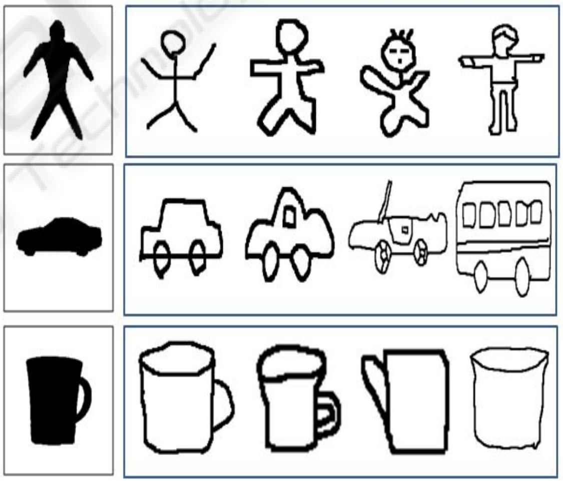
Abstract: Sketches drawn by users are one of the most intuitive forms of Human Computer Interaction. Users can easily express their intention by sketching simple hand-drawn lines.

In this paper, we consider the problem of target object detection and retrieval from a query by a sketch which is not in the database. Our novel approach consists of three steps: (1) Preprocessing to extract the skeletal features from a sketched query using size normalization, labelling, and binarization, (2) Skeletal feature extraction of query and data images in the space of diffusion tensor fields, and (3) Similarity measure using tensorial information between sketched query and database to retrieve the most similar target object in database.

Experiments are conducted to evaluate the performance of our methodology, which shows to be an efficient and mature retrieval system.

1 INTRODUCTION

There is a growing interest in new Human Computer Interaction techniques on tablet PCs, touch phones, and multi-touch screens (Han, 2006; Kim et al., 2007). Interaction using one or two fingers on these devices enable users to easily generate and retrieve multimedia items in databases. However, real HCI within these devices is still lacking in terms of understanding freehand input and low level interactive visualization of content. Among numerous interaction methodologies for detecting and retrieving the various images in database, we propose a freehand sketch which is the informal drawing of a shape using lines and curves as a natural interaction technique (Hassouna and Farg, 2007). Especially, the sketch image is very perceptive to human vision and a fast representation of important features of target objects for humans at all ages. Although a sketch is composed of few lines, it is a coarse but detailed picture including its key features. For instance, if we ask the people to draw objects like a human, a car, or a cup without any further information, most people intend to sketch the objects as shown in Figure 1. These sketches are passive and cannot be directly simulated or analyzed using computational engineering tools. Computers cannot understand the sketched images which have various viewpoints and non-textual information, even though users can easily understand their characteristics and classify the images into categories. Therefore, the sketched query image needs to be transformed into a computer-suited representation. Then robust features can be extracted that relate to the intention of the user.

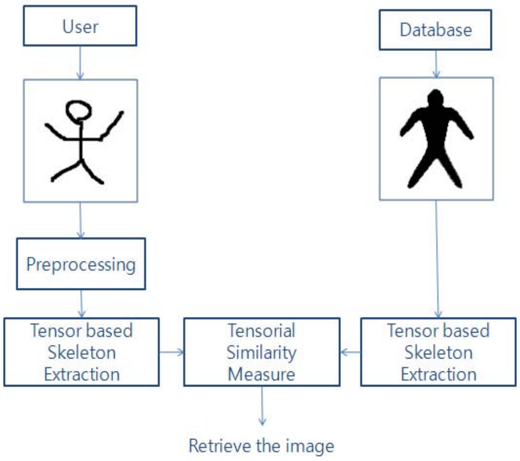


(a) Example images in the database (b) Sketches of target objects by several persons.

Figure 1: Comparison of the images <human, car, cup> in a database and in sketches. The sketched images are quickly drawn and easily created to visualize their shape, but they are different from user to user.

In this paper, we develop a query-by-sketch-based

Min Yoon S. and Kuijper A. (2010).
OBJECT RETRIEVAL BASED ON USER-DRAWN SKETCHES.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 292-299
DOI: 10.5220/0002818702920299


User
Database
Preprocessing
Tensor based Skeleton Extraction
Tensorial Similarity Measure
Tensor based Skeleton Extraction
Retrieve the image

Our approach is composed of tree steps: i) sketched query image transformation for size normalization, and binarization, ii) skeletal feature extraction using NGVF, and iii) similarity measuring to correctly detect and retrieve the images in database.

Figure 2: Structure of our proposed query-by-sketch based target object retrieval system.

target object detection and retrieval system by understanding the sketched query image, extracting the skeletal features of query image and database images, and measuring the similarity to correctly detect or retrieve the most similar object in database.

Figure 2 shows our methodology which we will explain in detail further on. First, the query-by-sketch image needs to be normalized to extract the efficient features by transforming the roughly sketched query image. We then extract the skeleton from binarized sketched image by using a novel topological analysis in the space of Normalized Gradient Vector Flow (NGVF) fields within a given image. Lastly, we retrieve the most similar object by measuring their tensorial feature characteristics.

Our proposed methodology has the following advantages compared to previous vector field based skeleton extraction and object retrieval technique:

(1) There is no need to determine a priori-information from images in the database or the sketch query image to detect the target objects.

(2) Our proposed methodology shows an improved skeleton extraction from a sketch image that includes a singular region.

(3) We can easily detect and retrieve the images in a database by its skeletal characteristics.

2 PREVIOUS WORK

Our query-by-sketch based image retrieval approach is based on some computer vision and image processing techniques. In this section, we briefly survey the previous work which is closely related to our approach, like sketch based image retrieval systems, skeletal feature extraction and its similarity measures, and diffusion tensor based object representations.

2.1 Sketch based Image Retrieval Systems

Research on image retrieval has been carried out in several fields such as computer vision, computational geometry, CAD/CAM, and molecular biology. Content based Image Retrieval (CBIR) systems allow users to search images in media databases in order to derive meaningful features, as well as to measure the dissimilarity of visual objects by distance functions (Maree07, Veltkamp01). Sketch based Image Retrieval (SBIR) systems have been developed as one of part of CBIR.

SBIR systems started from 2D image retrieval and were recently extended to 3D model retrieval and editing methods (Ip et al., 2001; Matusiak et al. 1998; Yang and Xiao, 2008). SBIR was developed to overcome the limitations of previously well-known approaches such as keyword or query-by-example based image retrieval. (Funkhouser et al., 2003) introduce a web-based search engine that has query images based on 2D or 3D sketches using a spherical harmonics shape descriptor. Hou et al. (Hou and Ramani, 2007) also presented a 3D model retrieval system using view-dependent 3D shape descriptors. The obvious advantage of this method is its ease for users to generate and retrieve the images. However, the boundary contours of each target object from different view directions, or information on incomplete shapes are needed and have to be prepared during a preprocessing phase. Fourier descriptors and Zernike Moments can also be used to match the sketched query image on retrieved images from a database (Hou and Ramani, 2007).

2.2 Skeleton Extraction

Previous skeleton computing methodologies can be roughly classified into three categories according to their approach: i) topological thinning, ii) distance transform based skeleton extraction, and iii) geometric modelling from the work of Blum (Blum, 1976), who defined the skeleton using the Medial Axis

Transform (MAT). However, existing skeleton extraction methodologies are still weak because of their high computational complexity, noise sensitivity, centeredness inside the underlying complex shape, and partial occlusion or artifacts within a singular region (Bai07,Bitter01,Chalecheale05,Ma03).

2.3 Diffusion Tensor Fields

Most CBIR and skeleton extraction systems are based on vector fields which are generated from a given image by different physical properties. Few work are investigated to extract the features in the space of diffusion tensor fields. Diffusion tensor fields (with a focus on symmetric, second order ones) are useful in many medical, mechanical, and physical applications such as fluid dynamics, meteorology, molecular dynamics, biology, astrophysics, mechanics, material science and earth science. (Basser et al., 1994) presented their work on diffusion tensor magnetic resonance imaging (DT-MRI). Using this new MRI modality, it was possible to qualify anisotropic properties of the imaged tissue by characterizing the water diffusion.

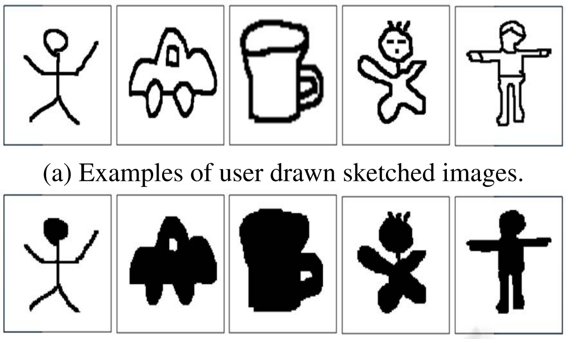
3 OUR APPROACH

In this section, we will explain our proposed approach for target object detection and retrieval using query-by-sketch image. It is composed of the three steps; i) sketched query based image transformation, ii) skeletal feature extraction using Normalized Gradient Vector Flow in the space of diffusion tensor fields, and iii) similarity measuring between a query image and an image dataset or target images.

3.1 Sketched Query Image Transformation

For sketch images which are drawn by users it is very difficult to understand their characteristics, because users sometimes omit important features or draw in detail with noisy lines. We therefore first transform the sketched query image to a simplified one in order to easily search and measure the similarity within a database. We transform the sketched image by using size normalization, labelling, and binarization methods.

The imported query-by-sketch image which is usually a rough and simple black hand-drawn figure with draft lines is firstly normalized to a size of 50 × 50 pixels. Then, we label the sketched query image (black / white). Separation of foreground and background of the sketched image is as follows: First, we recognize the labels which neighbor the background as foreground (that is, the black sketched parts). From the foreground labels, we iteratively check the label's connectivity with neighbor labels and separate background and foreground by adding background (white) labels to foreground when they are inside the object. Figure 3 shows the binarized query images that follow our proposed size normalization, labeling, and binarization scheme.



(a) Examples of user drawn sketched images.

(b) Transformed sketch images for skeletal feature extraction.

Figure 3: Sketch image transformation for skeleton extraction.

3.2 Skeletal Representation in Tensor Space

In this section, we will explain how we extract the skeletal features for measuring the similarity between query and database. It has been shown in literature that the skeletal feature based image retrieval systems are more efficient than shape based image retrieval systems (Zhang07). This is basically because skeletal features of a given image reduce data, while keeping its characteristics and being robust in singularities and under partial occlusion.

3.2.1 NGVF Fields from a Given Image

The Gradient Vector Flow (GVF) is a vector diffusion approach using Partial Differential Equations (PDEs). It converges towards the object boundary when it is very near to the boundary, but it varies smoothly over homogeneous image regions extending the image border. Originally, GVF fields were proposed to solve the problem of initialization and poor convergence to the boundaries of concave objects yielding a traditional snake form (Xu and Prince, 1998). The main advantage of GVF fields is that it is able to capture a snake over a long range and to force it into concave regions (Hassouna and Farg, 2007). Mathematically defined, the GVF is the vector field $\mathbf{v}$ that minimizes the following energy functional:

$$\varepsilon = \int\int \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + ||\nabla f||^2 ||\mathbf{v} - \nabla f||^2 dxdy. \tag{1}$$

Here $\mathbf{v} = [u(x,y), v(x,y)]$, and the initial value of $\mathbf{v(x,y)}$ is determined by $\nabla f(x,y)$. $\nabla f(x,y)$ is the gradient image derived from a given image. μ is a regularization parameter that is to be set based on noise present in image. Minimizing this energy will force $\mathbf{v}(x,y)$ to be nearly equal to the gradient of the edge map, where $||\nabla f(x,y)||$ is large. Nevertheless, the general GVF method cannot efficiently extract the medial axis, since a weak vector has very little impact on its neighbors that have much stronger magnitudes.

A Normalized Gradient Vector Flow (NGVF) can tremendously affect a strong vector, both in magnitude and in orientation by normalizing the vectors over the image domain during each diffusion iteration (Yu and Bajaj, 2002).

The traditional GVF has difficulty preventing the vectors on the boundary from being significantly influenced by the nearby boundaries and thus cause a problem such that the "snake" may move out of the boundary gap. NGVF avoids this problem by normalizing its direction and magnitude (Yoon et al, 2009).

3.2.2 Skeleton Extraction with Diffusion Tensor Field

Previous skeleton extraction and structural analysis techniques are computed in a vector fields and their topologies are obtained by locating critical points and displaying the set of their connecting streamlines. When it comes to representing and analyzing the directions of NGVF from a given image, tensor fields provides more information than vector fields. The tensor field is defined as a symmetric, second-order matrix at each point as follows:

$$T(\bar{x}) = \begin{pmatrix} T_{xx}(x,y) & T_{xx}(x,y) \\ T_{xy}(x,y) & T_{yy}(x,y) \end{pmatrix} \tag{2}$$

$T(\bar{x})$ is can be expressed using its eigensystem as

$$T(\bar{x})\bar{e}_i(\bar{x}) = \lambda_i(\bar{x})\bar{e}_i(\bar{x}), \tag{3}$$

where i=1,2, $\lambda_i(\bar{x})$ are the eigenvalues of $T(\bar{x})$, and $\bar{e}_i(\bar{x})$ define the eigenvectors. Details of the skeleton extraction and ellipsoidal representation from a binalized target object can be found in e.g. (Yoon and Graf, 2009). The scale and orientation of each ellipse at each pixel are determined by its eigenvalues and eigenvectors. Figure 4 is a conceptual representation of an edge image with extracted eigenvalues and eigenvectors. The ratio between eigenvalues determines the shape of the ellipsoids, their sum defines their scale and its principal eigenvector direction defines the rotation of the ellipsoid. Using these tensorial elements in each pixel, we can reduce the pixels of the images within a database and the query image to an ellipsoidal representation.

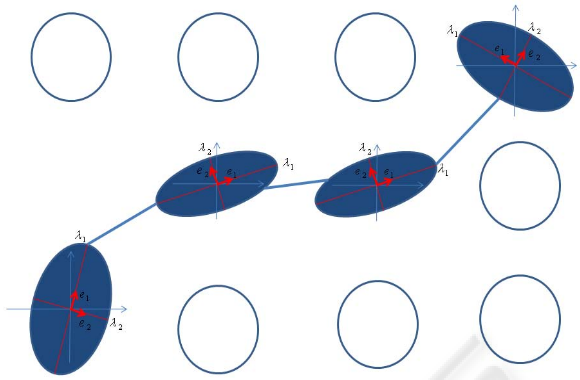


Figure 4: Each pixel of image can be represented as extracted eigenvalues and eigenvectors which are extracted from tensorial properties. In a given image, we separate the edge and other pixels using edge detection. Edge pixels are expressed as ellipsoid with their eigenvalues and eigenvectors and other pixels are shown as circles.

Figure 6a shows the skeletons of images which are extracted by analyzing the tensorial features and Figure 6b displays the skeletal characteristics with ellipsoidal representation using the eigenvectors and eigenvalues which are calculated from the tensor elements.

3.3 Similarity Measure between Sketch and Database Images

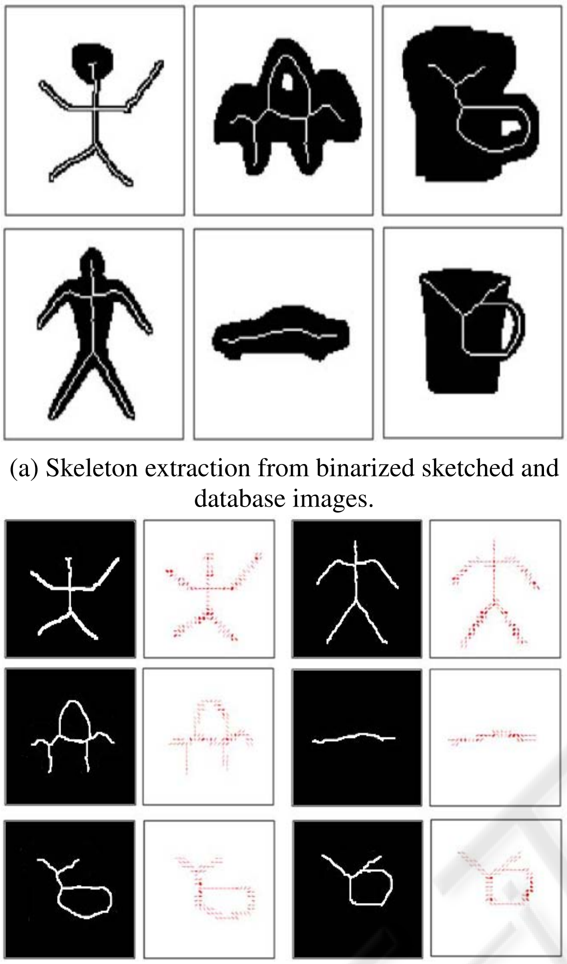
Before measuring the similarity, we separate the extracted skeleton into several branches. Thus, we can express the skeletal features from the sketched query image, I_i, and the image in database I_j, as

$$F_i = \{f_1^i, f_2^i, ..., f_{p-1}^i, f_p^i\}$$

and

$$F_j = \{f_1^j, f_2^j, ..., f_{q-1}^j, f_q^j\},$$

where p and q are the number of branches in I_i and I_j, respectively. Figure 6 shows the F_i and F_j using example images and their skeletal features. The scale and rotation of each branch are represented its averaged eigenvalues and eigenvectors. The tensorial elements and eigen-features of F_i and F_j are used to measure the similarity measure between images.



(a) Skeleton extraction from binarized sketched and database images.

(b) Ellipsoidal representation of the extracted skeleton of the target objects.

Figure 5: Skeleton extraction and the ellipsoidal representation of target objects using their tensorial properties.

The score matrix of many to many matching of skeletal features can be expressed as:

$$Score_{matrix} = \frac{1}{N} \begin{pmatrix} s^{ij}_{11} & s^{ij}_{12} & \cdots & s^{ij}_{1q} \\ s^{ij}_{21} & s^{ij}_{22} & \cdots & s^{ij}_{2q} \\ \vdots & \cdots & \cdots & \vdots \\ s^{ij}_{p1} & s^{ij}_{p2} & \cdots & s^{ij}_{pq} \end{pmatrix}, \qquad (4)$$

where N is a normalization factor. Since the score matrix in tensorial analysis is the proper choice of the similarity measure to be used, we will explain in the following how to extract the s^{ij}.

Given two tensorial elements, $\mathbf{T}_i$, and $\mathbf{T}_j$ the most simple comparison between two tensor quantities is the tensor dot product (Delarcelle and Hesselink, 1994):

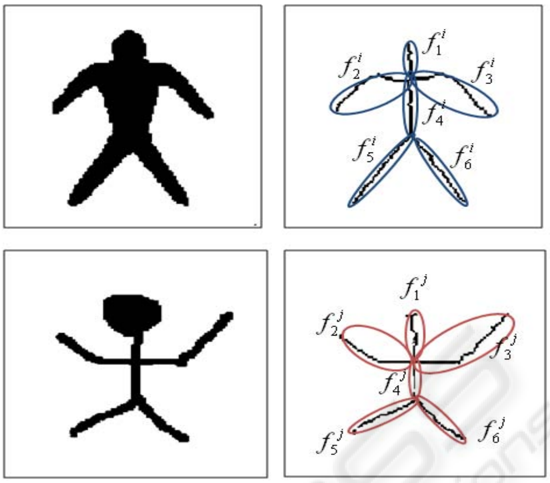


Figure 6: Ellipsoidal representation of branches by using averaged eigenvalues and eigenvectors. The similarity measure between image is determined by characteristic of F_i and F_j.

$$d_1(\mathbf{T}_i,\mathbf{T}_j) = \lambda_{1,i}\lambda_{1,j}(e_{1,i}\cdot e_{1,j})^2 + \lambda_{2,i}\lambda_{2,j}(e_{2,i}\cdot e_{2,j})^2. \qquad (5)$$

One such example is the tensor Euclidean distance obtained by using the Frobenius norm (Alexander99).

Due to its simplicity, tensor Euclidean distance has been used extensively in DTI restoration.

$$d_2(\mathbf{T}_i,\mathbf{T}_j) = \sqrt{Trace((\mathbf{T}_i - \mathbf{T}_j)^2)}. \qquad (6)$$

From various similarity measure methods, we measure the similarity measure as the multiplication of Eqs. (5-6):

$$s^{ij} = d_1(\mathbf{T}_i,\mathbf{T}_j)\cdot d_2(\mathbf{T}_i,\mathbf{T}_j). \qquad (7)$$

This combines the properties of the two similarity measures, namely the difference in scale and the difference in angle of elliptical tensorial elements.

The score matrix s^{ij} is merged to each labeled branch and we can recalculate the similarity measure between the labels. The final score is determined by combining the minimum similarity values from each branch.

4 EXPERIMENTS

We performed experiments in order to extract the skeleton of query by sketch images using our proposed approach. As we explained in section 3, the imported sketch images are converted to a binary format with a 50×50 pixel size to correctly detect and retrieve the target objects in the database. Afterwards,

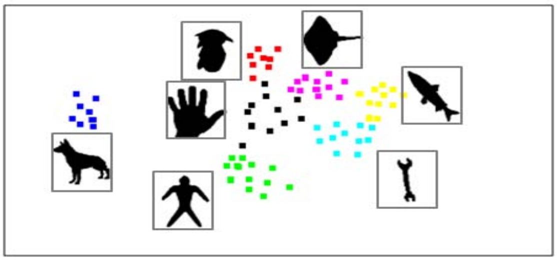


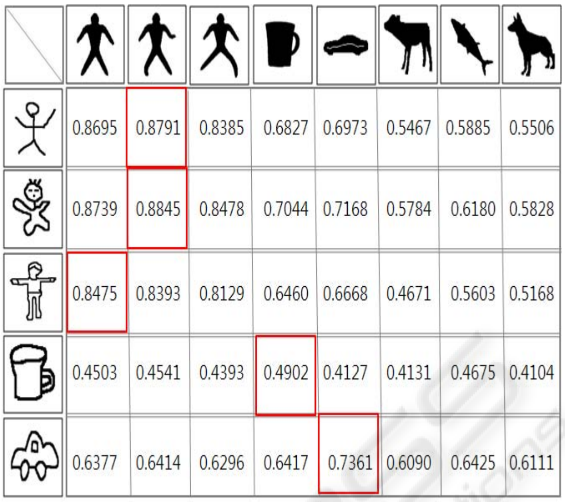
Figure 7: Classification of target objects in LEMS 99 image dataset using our proposed similarity measure. The images are the representative image in each class which is painted by different color.

we measure the similarity between a query image and the database images.

Before search the most similar target object in the LEMS 99 image dataset, we classify the target objects using our proposed skeletal feature extraction and similarity measure methodology shown in Figure 7. One can see that the two degree of freedom correctly clustered the images. Figure 8 shows the sketched and some retrieved objects with their similarity measure in a matrix showing the query images and some images from the LEMS 99 image dataset. The red box shows the highest score of a sketched query image and the found image from the dataset. We only showed a selection with the best scores form the LEMS data base for visualization purposes.

Figure 9 shows the extracted silhouettes given one sketched query image. The test image is composed of various human body objects and others including cars. We first labeled each object and measured the similarity between the query image and the various target objects in image. The target objects which are painted with a red box are the detected objects from the simply sketched query image. Figure 9 also shows that our approach is very robust in partial occlusion and pose transform of human body object. Obviously, adding more sketch input images with different topological properties would increase the number of positive detections - like persons with both arms and legs occluded / coinciding, etc. However, given only one general query image we were already able to detect most of the humans.

We also tested our proposed method on various users to see how our system performs on input query images that differ from user to user. We asked user to draw a chair and searched for the closed hits in a chair data base. Figure 10 shows the retrieved "chair" images from 5 different users. As we asked them to draw the chair without specifying further information, the drawn sketch images differ from case to case. The score value and ordering is different according to the similarity measure. Especially, the best scores of the retrieved images for user 4 are lower than the others retrieved top ranked images, because the database in the chair class does not have similar pictures.



	0.8695	0.8791	0.8385	0.6827	0.6973	0.5467	0.5885	0.5506
	0.8739	0.8845	0.8478	0.7044	0.7168	0.5784	0.6180	0.5828
	0.8475	0.8393	0.8129	0.6460	0.6668	0.4671	0.5603	0.5168
	0.4503	0.4541	0.4393	0.4902	0.4127	0.4131	0.4675	0.4104
	0.6377	0.6414	0.6296	0.6417	0.7361	0.6090	0.6425	0.6111

Figure 8: Similarity measure evaluation test using one sketched image. The red box highlights the highest scored target object in the database.

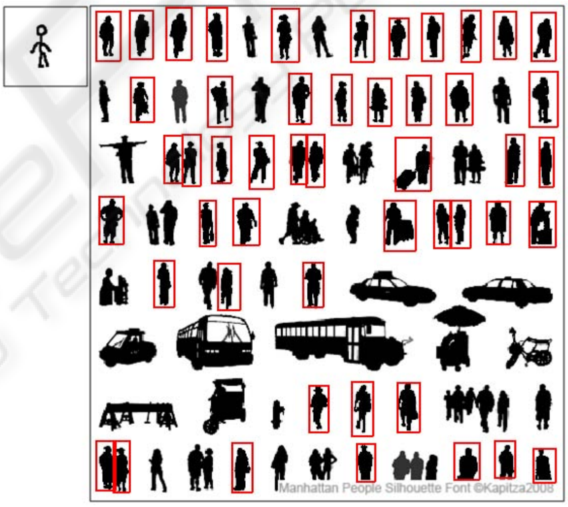


Figure 9: Extraction of matched silhouettes from one given image. The red box highlights target objects that are detected from this given sketched query image.

5 CONCLUSIONS AND FUTURE WORK

The query-by-sketch based image retrieval system is a very efficient method to express users' intension for Human Computer Interaction. The performance

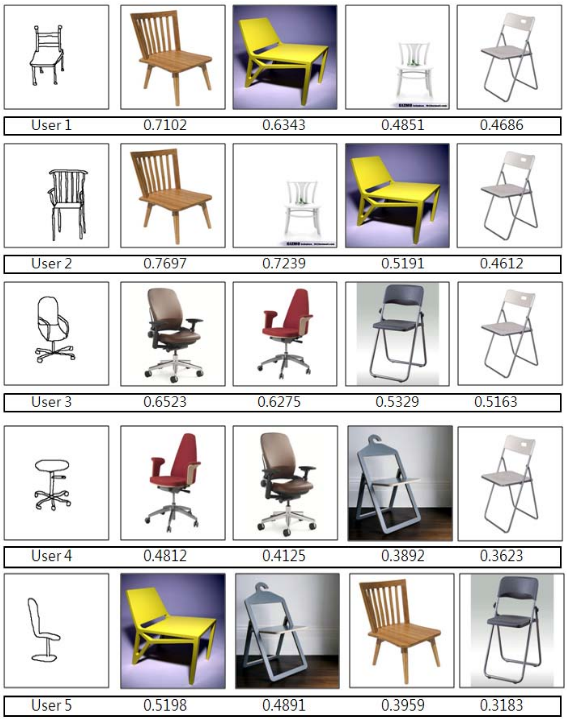


Figure 10: Image retrieval for a sketched ”chair” image from various users.

of SBIR systems is very dependent on the nature of complex image data, on the extraction of meaningful features from complex images, and on the similarity measure determined by a roughly sketched image.

In this paper, we have presented our approach for extraction of tensorial features and the measurement of similarities, as well as enhanced image classification techniques. The essential idea is based on an analysis of a tensor topology in order to extract the ellipsoidal characteristics of features. We have also shown that our methodology is very efficient in retrieving the most similar images from a large repository in a short time. It is scalable due to the addition of new images into the database.

Our proposed sketch based image retrieval system is not limited to 2D image search and retrieval. We are currently working on extending our methodology to understanding of 2D/3D motions of target objects. Especially, we are focusing on a 3D structural analysis of objects in the space of tensor fields.

REFERENCES

Alexander, D., Gee, J., and Bajcsy, R. (1999). Similarity measures for matching diffusion tensor images. In *Proceedings of BMVC.*

Bai, X., Latecki, L. J., and Liu, W. Y. (2007). Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution. *IEEE trans. on PAMI.*

Basser, P. J., Mattiello, J., and Le Bihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys Journal.*

Bitter, I., Kaufman, A. E., and Sato, M. (2001). Penalized-distance volumetric skeleton algorithm. *IEEE trans. on Visualization and Computer Graphics.*

Blum, H. (1976). A transformation of extracting new descriptions of shape. *Models for Perception of Speech and Visual Forum.*

Chalecheale, A., Naghdy, G., and Mertins, A. (2005). Sketch-Based Image Matching Using Angular Partitioning. *IEEE trans. on Systems, man, and cybernetics.*

Delarcelle, T., and Hesselink, L. (1994). The topology of symmetric, second-order tensor fields. In *Proceedings of IEEE Visualization.*

Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., and Jacobs, D. (2003). A search engine for 3D model *ACM trans. on Graphics.*

Han, J. Y. (2006). Multi-touch interaction wall. In *SIGGRAPH '06: ACM SAIGGRAPH Emerging technologies.*

Hassouna, M. S., and Farg, A. A. (2007). On the extraction of Curve skeletons using Gradient Vector Flow. In *Proceedings of ICCV.*

Hou S., and Ramani, K. (2007). Classifier combination for sketch-based 3D part retrieval. *Journal of Computers Graphics.*

Ip, H. H. S., Cheng, A. K. Y., Wong, W. Y. F., and Feng, J. (2001). Affine-invariant sketch-based retrieval of images. In*Proceedings of ICCG.*

Kim, J., Park, J., Kim, H., and Lee, C. (2007). HCI using multi-touch tabletop display. *Communications, Computers, and Signal Processing, PacRim.*

Ma, W. C., Wu, F. C., and Ouhyoung, M. (2003). Skeleton extraction of 3D objects with radial basis functions In *Proceedings of the Shape Modeling.*

Maree, R., Geurts, P., and Wehenkel, L. (2007). Content-Based Image Retrieval by Indexing Random Subwindows with Randomized Trees. In *Proceedings of ACCV.*

Matusiak, S., Daoudi, M., Blu, T., and Avaro, O. (1998). Sketch-based images database retrieval. In *Proceedings of Int. Workshop Adv. Multimedia Information System.*

Vasconcelos, N., and Lippman, A. (2000). A Probabilistic Architecture for Content-based Image Retrieval. In *Proceedings of CVPR.*

Veltkamp, R. C., Burkhardt, H., and Kriegel, H. P. (2001). State-of-the Art in Content Based Image and Video Retrieval. *Kluwer Adcademic Publshers.*

Xu, C., and Prince, J. L. (1998). Snakes, shapes, and gradient vector flow. *IEEE trans. on Image processing.*

Yang, G., and Xiao, Y. (2008). A robust similarity measure method in CBIR system. In *Proceedings of Congress on ISP.*

Yoon, S. M., and Graf, H. (2009). Automatic skeleton extraction and splitting of target objects. *IEEE ICIP.*

Yoon, S. M., Malerczyk, C., and Graf, H. (2009). 3D skeleton extraction from volume data based on Normalized Gradient Vector Flow. In *Proceedings of WSCG.*

Yu, Z., and Bajaj, R. (2002). Normalized gradient vector diffusion and image segmentation. In *Proceedings of ECCV.*

Zhang, E., Hays, J., and Turk, G. (2007). Interactive tensor field design and visualization on surfaces. *IEEE trans. on Visualization and Computer Graphics.*