

June 2002

Face recognition using a similarity-based measure in image database for crime investigation

V. Srisarkun
University of Wollongong

J. Cooper
University of Wollongong

Follow this and additional works at: <https://ro.uow.edu.au/infopapers>



Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Srisarkun, V. and Cooper, J.: Face recognition using a similarity-based measure in image database for crime investigation 2002.
<https://ro.uow.edu.au/infopapers/129>

Face recognition using a similarity-based measure in image database for crime investigation

Abstract

A method to handle searching a face image database (FID) is proposed to support police officers when searching criminal records from a central registration database system (CRDS). The proposed method assumes that each FID consists of a fixable object and object correlation. The proposed method employs a database search, so that all images with a similarity-based measure are retrieved. Consequently, the proposed method is much faster than sequential searching, especially when an additional set of attributes, like scar, is defined. Moreover it requires less storage space.

Keywords

face recognition, image retrieval, police data processing, query formulation, visual databases

Disciplines

Physical Sciences and Mathematics

Publication Details

This paper originally appeared as: Srisarkun, V & Cooper, J, Face recognition using a similarity-based measure in image database for crime investigation, 4th EURASIP-IEEE Region 8 International Symposium on VIPromCom Video/Image Processing and Multimedia Communications, 16-19 June 2002, 87-92. Copyright IEEE 2002.

FACE RECOGNITION USING A SIMILARITY-BASED MEASURE IN IMAGE DATABASE FOR CRIME INVESTIGATION

Vilasinee Srisarkun and Joan Cooper

Faculty of Informatics,
University of Wollongong, NSW 2522, Australia.
[vs12@uow.edu.au]

Abstract

In this paper a method to handle searching for face image database (FID) is proposed to support police officers when searching criminal records from a Central Registration Database Systems (CRDS). The proposed method assumes that each FID consists of a fixable object and object correlation. The proposed method employs a database search, so that all images with similarity-based measure will be retrieved. Consequently, the proposed method is much faster than sequential searching especially when an additional set of attribute like scar is defined. Moreover it requires less storage space.

Key words: *face image database, similarity, simulation, correlation and object relation.*

1. INTRODUCTION

To support queries by image content in a Face Image Database (FID), all images must be analyzed prior to storage, so those attributes can be extracted and stored in the database separately. These descriptions are then used to search the FID and to determine which images satisfy the query selection criteria. Matching between these attributes is a well-known method so called self-similar has been studied extensively in [2], [3], [5] and [6]. The effectiveness of a FID system ultimately depends on the image representations, the types of image queries, and the efficiency of searching technique and the traffic congestion of network when accessing to the CRDS [4].

The user may specify several objects with complex shapes and interrelationships and may ask for all images containing self-similar. The term self-similar means that at every time the retrieved images from FID are resemblance to what the users query. Then database searching must distribute the self-similar ones so that all images ranging up to a specified correlation will be retrieved. This paper concentrates on analysis of query for various correlation parameters in FID searching. Self-similar FID can be recognized in multifirmity: A density of original face image content, comparison cost that is a similarity-based measurement [1] and a variance of sample means that sample size is essential with the time-scale comparison.

In sequential searching, the algorithm compares the stored face image in central registration database (CRDS) to the query image pixel by pixel, starting from the first pixel until all pixels are covered. The perfect-matched image will only be retrieved, when all pixel values of stored image are matched. As the sequential searching compares the pixel value of the two images pixel by pixel at a time, and it is evident that an image comprises of great

number of pixels; the sequential searching time depends on the image size. The larger size is the longer comparing time will be. For example, if an image of 90*120 pixels with 256 monochrome scales will be used then one image produces totally 10,800 pixels to compare.

The methods proposed in this paper require few attributes to represent the content of face image. The attribute number is adequate to retrieve the similar face images from the database. The space required for attributes storage is insignificant and search time will be considerably reduced. The performance of proposed method will apparently improve with 100-database size. However, in this paper the performance will be measured by comparing a larger size of database, which is up to 10,000 with a specification in scar refinement [7].

2. PROPOSED METHOD

2.1 Object Number

Face image consists of objects, which are identified prior to their storage. Each object has the fixed object number starting from zero (e.g., 0 is the relation between face and right eyebrow) as shown in Table 1. Attributes (e.g., size at 0 degree, reference angle and distance between those two reference objects) are also included. Each object stored in the database will be compared to each object of the key image (attribute by attribute). In fact the comparison may or may not emanate a perfect match. But the image with a satisfied degree of self-similar is retrieved.

2.2 Image Conversion

The face image is segmented into closed contours corresponding to dominate image objects. Each object contains object number and attributes. The object numbers of a face image are shown in Table I. Moreover, since each object in the face image is non-roundness object, each object includes a number of attributes as follows: Size at 0 degree, Size at 90 degree, Size at 180 degree, Size at 270 degree. The unit of these attributes is measured in millimeter. In order to reduce the number of object and calculation complexity, we will refer to a center point for all objects as shown in Table 1. For example, let n be the number of objects of an image. If the center point is reference then object number is $(n-1)$. If no reference point is needed then the number of object will be $\sum_{i=1}^{n-1} n-i$.

Table I. Object Number

Objects (reference from center of the face image)	Object Number
Center of the Face	0
Right Eyebrow	1
Left Eyebrow	2
Right Eye	3
Left Eye	4
Right Ear	5
Left Ear	6
Nose	7
Mouth	8
Scar	9

Each object contains 3 attributes; object number, distance, and angular direction to the reference point. The distance is measured in millimeter while the direction is measured in degree. In this research, the objects and object correlation are estimated prior to the storing. The amount includes size, distance, and angle. Searching for stored images in the database requires that all images with positive correlation will be retrieved. Comparing to sequential searching, our searching will retrieve the self-similar images from the database faster. The likeness of the output images depends on the degree of self-similarity as shown in figure 2.

3. EXPERIMENTS

3.1 Object Correlation Estimation

In order to perform self-similarity searching, which is faster in searching the image from the database and obtains the likeness image, the self-similarity, searching method for FID is applicable. The face image, which is taken from the Central Registration Database, was enlarged 6 times then converted into objects and object number. The attribute values of each object and object numbers are then identified. Next, all data is stored in the database. The steps used to identify the objects, object number, and attribute values are as follows.

1. For each face image, there are 10 objects – Face center, Right Eyebrow, Left Eyebrow, Right Eye, Left Eye, Right Ear, Left Ear, Nose, Mouth, and Scar.
2. For each object, the center of the object is identified.
3. Then, for each object, the size at 0 degree, size at 90 degree, size at 180 degree, and size at 270 are measured from the center of the object. If any objects do not exist in the face image, these attribute values are set to a default value of zero.
5. For all face images, there are 9 objects number in which Face object is the reference object – Face Center (FC) towards Right Eyebrow, FC towards Left Eyebrow, FC towards Right Eye, FC towards Left Eye, FC towards Right Ear, FC towards Left Ear, FC towards Nose, FC towards Mouth, and FC towards Scar.
6. Then, each attribute of the object is identified. The distance between 2 objects is measured in millimeter from the center of these 2 objects.

The angle between 2 objects is measured in the horizontal direction of the line connecting between these 2 object centers. If any object does not exist in the face image, these attribute values are set to zero. Finally, all contents of the face image are stored into the database.

In comparison between each image in the database and key image, measurement of well-matching is necessary. The cost of matching between the two objects will be defined as degree of similarity d which is the cost percentage of well-matching of the objects or object correlation between associated objects. Let k be the number of attributes of an object or object correlation. Let q be an attribute value of the object of the key image and r is an attribute value of the object of stored image. The absolute values of q and r are computed as follows:

$$q_i = |q_i| \text{ and} \quad (1)$$

$$r_i = |r_i| \quad (2)$$

where $i = 0, 1, 2, \dots, k-1$

The cost of each attribute between 2 objects is defined as:

$$\text{cost} = |q_i - r_i| \quad (3)$$

where $i = 0, 1, 2, \dots, k-1$

Then, cost percentage is defined as:

$$\text{cost_percentage} = \frac{|q_i - r_i|}{\max(q, r)} * 100 \quad (4)$$

For example, the *cost_percentage* of face object and object correlation between face object and right eyebrow of an image is computed as follows:

Size at 0 degree

$$\begin{aligned} \text{cost_percentage} &= \frac{|51 - 52|}{52} * 100 \\ &= 1.92 \% \end{aligned} \quad (5)$$

Similarly, size at 90 degree = 1.23 %

Size at 180 degree = 0.96 %

Size at 270 degree = 1.22 %

Distance = 4.84 %

Angle = 0.67 %

Similarity searching in the database of stored images requires that all images within the pre-specified degree of similarity d must be retrieved. That is, all images with the following condition are retrieved.

$$\text{cost_percentage} \leq d \quad (6)$$

In comparison with sequential searching, similarity searching will retrieve the similar images from the database with faster search time. The similarity of the output images depends on the degree of similarity. The lower similarity, the closer image will be. However, in sequential searching, it is taking longer time to find the similar images in the database.

3.2 Query By Example

Query program written in C language is implemented. To check the efficiency of the proposed method, a sample up to 5,000 face images is employed in order to find the images that including scar or birthmark located on the right cheek of the face. However, large data sets of face images are complicate to identify attributes, object correlation and so on. The software that automatically converts the face image into the data is currently being developed. PC with PentiumIII 450MHz is used for the experiment. All databases had been stored in the hard disk. The experiments have been divided into several categories with a different number of images stored in the database as of 1000, 2000, 3000, 4000, and 5,000. For each category, as specified in section 4, five images in the database with cost = 0% are randomly used for computing the average time of sequential searching and similarity searching method.

4. RESULTS AND ANALYSIS

The experiment of sequential searching and self-similarity searching by using different stored images (1000, 2000, 3000, 4,000 and 5,000) is queried by example. The search is performed five times and eventually is computed as an average search time. The following tables show the average search time in each group.

**Table II: Average Search Time (sec) of Sequential Search versus Self-Similarity Search
Experimental Results**

Number of Image = 1,000

TRIAL	1st	2 nd	3rd	4th	5th	AVG
Sequential	42.3	41.6	41.2	41.8	42.3	41.9
Self-Similar	5.8	6.1	6.4	6.6	6.1	6.2

Number of Images = 5,000

TRIAL	1st	2 nd	3rd	4th	5th	AVG
Sequential	209	211.5	208	210	209	209.5
Self-Similar	29	28	28	32	33	30

Table II indicates the average response time for sequential search and self-similarity search with a different number of faces images in the database. Depending on the total number of image, the response time for both sequential search and self-similarity search increases. However in self-similarity search, the response time is gradually increasing compared to ones as of sequential search time is seven times. Figure 1 shows samples of the result images that including scar from the Central Registration Database System after the key image was specified.

5. CONCLUSIONS

The self-similarity searching method proposed in this paper could be applied to significantly reduce the crime records searching time as the searching result emanate the similar faces of a suspect for further decision-making in the crime investigation process. Previous research shown that by using a self-similarity searching method in comparison with sequential search when searching in image database is resulting in a faster search time and not to mention that a space required for the image storage is also reduced. However, the present research shown that a closer likeness image from the database is retrieved much faster when an additional set of attribute like scar or birthmark is specified as shown in figure 3.

REFERENCES

- [1] Lawrence, Yianilos and Cox "Face Recognition using Mixture-Distance and Raw Images," *IEEE International Conference on Systems, Man, and Cybernetics*, 1997, 2016-2021.
- [2] Squire, "Learning a Similarity-Based distance measure for Image Database Organization from Human Partitionings of an Image Set," *The 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*, 1998, 88-93.
- [3] Santini, "Similarity Matching" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(4), 1999, 34-40.
- [4] V. Srisarkun, et.al, "Traffic Shaping for Frame Relay over ATM Network," *Proceeding of the 13th IASTED International conference on Parallel and Distributed Computing and Systems*, 2001, 469-473.
- [5] V. Srisarkun, et.al, " A Model for Self-Similar Searching in Face Image Data Processing," *the Sixth INFORMS Conference on Information Systems & Technology*, 2001, 389-397.
- [6] V. Srisarkun, et.al, " Self-Similar Searching in Image Database for Crime Investigation," *the twentieth IASTED International Conference on Applied Informatics*, 2001.
- [7] V. Srisarkun, et.al, " A model for Self-Similar Search in Image Database with Scars," *Submitted for publication in The 10th European Conference on Information Systems (ECIS'02)*, 2002.

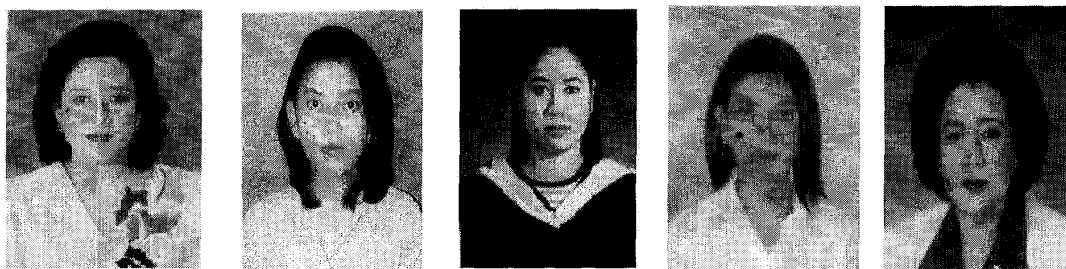


Figure 1 : Sample of the result images that including scar from the CRDS

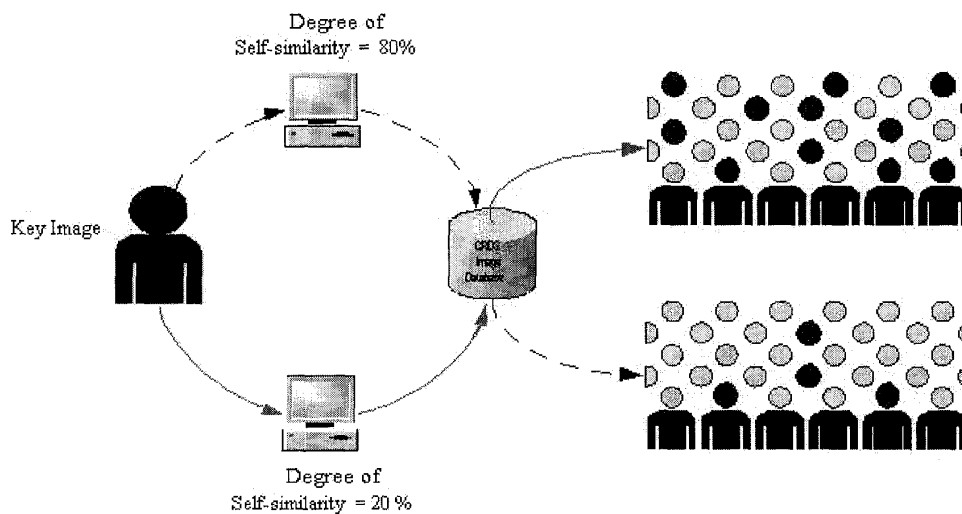


Figure 2: The likeness of the output images depends on the degree of self-similarity

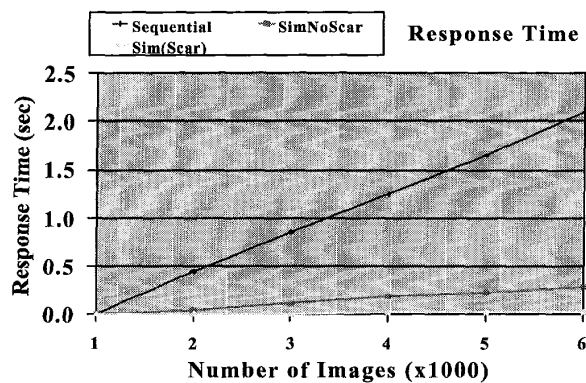


Figure 3 : Response Time of Sequential Search, Similarity Search with and without scar specification