DECREASING VOLUME OF FACE IMAGES DATABASE AND EFFICIENT FACE DETECTION ALGORITHM

Grigor A. Poghosyan and Hakob G. Sarukhanyan

Abstract: *As one of the most successful applications of image analysis and understanding, face recognition has recently gained significant attention. Over the last ten years or so, it has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. Biometric face recognition, otherwise known as Automatic Face Recognition (AFR), is a particularly attractive biometric approach, since it focuses on the same identifier that humans use primarily to distinguish one person from another: their "faces". One of its main goals is the understanding of the complex human visual system and the knowledge of how humans represent faces in order to discriminate different identities with high accuracy.*

Human face and facial feature detection have attracted a lot of attention because of their wide applications, such as face recognition, face image database management and human-computer interaction. So it is of interest to develop a fast and robust algorithm to detect the human face and facial features. This paper describes a visual object detection framework that is capable of processing images extremely rapidly while achieving high detection rates.

Keywords: *Haar-like features, Integral Images, LEM of image- Line Edge Map, Mask $n \times n$ size*

Introduction

As the necessity for higher levels of security rises, technology is bound to swell to fulfill these needs. Any new creation, enterprise, or development should be uncomplicated and acceptable for end users in order to spread worldwide. This strong demand for user-friendly systems which can secure our assets and protect our privacy without losing our identity in a sea of numbers, grabbed the attention and studies of scientists toward what's called biometrics.

Biometrics is the emerging area of bioengineering; it is the automated method of recognizing person based on a physiological or behavioral characteristic. There exist several biometric systems such as signature, finger prints, voice, iris, retina, hand geometry, ear geometry, and face. Among these systems, facial recognition appears to be one of the most universal, collectable, and accessible systems.

Biometric face recognition, otherwise known as Automatic Face Recognition (AFR), is a particularly attractive biometric approach, since it focuses on the same identifier that humans use primarily to distinguish one person from another: their "faces". One of its main goals is the understanding of the complex human visual system and the knowledge of how humans represent faces in order to discriminate different identities with high accuracy.

The face recognition problem can be divided into two main stages:

- face verification (or authentication)
- face identification (or recognition)

The detection stage is the first stage; it includes identifying and locating a face in an image.

Face detection has been regarded as the most complex and challenging problem in the field of computer vision, due to the large intra-class variations caused by the changes in facial appearance, lighting, and expression. Such variations result in the face distribution to be highly nonlinear and complex in any space which is linear to the original image space. Moreover, in the applications of real life surveillance and biometric, the camera limitations and pose variations make the distribution of human faces in feature space more dispersed and complicated than that of frontal faces. It further complicates the problem of robust face detection.

Face detection techniques have been researched for years and much progress has been proposed in literature. Most of the face detection methods focus on detecting frontal faces with good lighting conditions. According to Yang's survey [Yang, 1996], these methods can be categorized into four types: knowledge-based, feature invariant, template matching and appearance-based.

Knowledge-based methods use human-coded rules to model facial features, such as two symmetric eyes, a nose in the middle and a mouth underneath the nose.

Feature invariant methods try to find facial features which are invariant to pose, lighting condition or rotation. Skin colors, edges and shapes fall into this category.

Template matching methods calculate the correlation between a test image and pre-selected facial templates.

Appearance-based, adopts machine learning techniques to extract discriminative features from a pre-labeled training set. The Eigenface method is the most fundamental method in this category. Recently proposed face detection algorithms such as support vector machines, neural networks [Raul, 1996] statistical classifiers and AdaBoost-based [Tolba, 2006] face detection also belong to this class.

Any of the methods can involve color segmentation, pattern matching, statistical analysis and complex transforms, where the common goal is classification with least amount of error. Bounds on the classification accuracy change from method to method yet the best techniques are found in areas where the models or rules for classification are dynamic and produced from machine

learning processes.

The recognition stage is the second stage; it includes feature extraction, where important information for discrimination is saved, and the matching, where the recognition result is given with the aid of a face database. Among the different biometric techniques facial recognition may not be the most reliable and efficient but it has several advantages over the others: it is natural, easy to use and does not require aid from the test subject.

Because the face detection and recognition database is a collection of images and automatic face recognition system should work with these images, which can hold large volumes of computer memory that is way it's necessary to investigate and develop a method / tool for optimal using volume of computer memory (that decrease image database volume) and implement quick face detection within database.

In this paper investigations and study of certain methods that helps us to develop and implemented method/tool , which decreases volume of face images database and use these database for face detection and recognition .To achieve the above mentioned goal we are using image LEM (Line Edge Map) [Tolba, 2006] of image and Viola Jones method [Viola and Jones, 2001]. As quantity of images in recognition database (n) is not infinite and we need define format for images (*.bmp, *.jpg), in this work we've selected n=40 and *.jpg format [Face Database]. As object detection and recognition algorithm we are using Haar-like features [Viola and Jones, 2001]. Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector.

This paper is organized as follows. Section 2 and 3 presents the Line Edge Map (LEM) of image and Viola Jones method (Haar like features/ classifier cascades). Section 4 presents using image LEM in Viola Jones method as input image, which will decreases volume of our face images database. Section 5 shows the experimental results of frontal face detection using constructed face images database (LEM) and Haar-like features and conclusions follows in Section 6.

Line Edge Map (LEM)

Edge information is a useful object representation feature that is insensitive to illumination changes to certain extent. Though the edge map is widely used in various patterns recognition fields, it has been neglected in face recognition except in recent work reported in [Takacs, 1998].

Edge images of objects could be used for object recognition and to achieve similar accuracy as gray-level pictures. In [Takacs, 1998] made use of edge maps to measure the similarity of face images and 92% accuracy was achieved. Takacs [Takacs, 1998] argued that process of face recognition might start at a much earlier stage and edge images can be used for the recognition of faces without the involvement of high-level cognitive functions.

A LEM approach, proposed by [Gao, 2002], extracts lines from a face edge map as features. This approach can be considered as a combination of template matching and geometrical feature matching. The LEM approach not only possesses the advantages of feature-based approaches, such as invariance to illumination and low memory requirement, but also has the advantage of high recognition performance of template matching.LEM integrates the structural information with spatial information of a face image by grouping pixels of face edge map to line segments. After thinning the edge map, a polygonal line fitting process [Leung, 1990] is applied to generate the LEM of a face. An example of LEM is illustrated in Fig. 1.

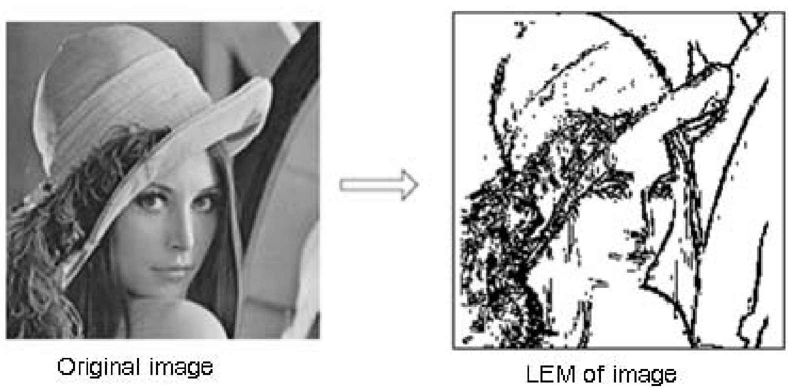

Figure 1. Used 8x8 mask for getting image LEM

Experiments on frontal faces under controlled /ideal conditions indicate that the proposed LEM is consistently superior to edge map. LEM correctly identify 100% and 96.43% of the input frontal faces on face databases [Bern, 2002, Purdue Univ. Face Database], respectively.

Viola Jones method (Haar like features/ classifier cascades)

Image sensors have become more significant in the information age with the advent of commodity multi-media capture devices such as digital cameras, webcams and camera phones. The data from these media sources (whether they are still images or video) is reaching the stage where manual processing and archiving is becoming impossible. It is now possible to process these images and videos for some applications in near real-time, using motion detection and face tracking for security systems for example. However there are still many challenges including the ability to recognize and track objects at arbitrary rotations.

Haar-like features have been used successfully in image sensors for face tracking and classification problems [Konstantinos, 2008, Duda, 2001], however other problems such as hand tracking [Shan, 2004] have not been so successful. Historically, working with only image intensities ((i.e., the RGB pixel values and every pixel of image) made the task of feature calculation computationally expensive. In [Papageorgiou, 1998] discussed working with an alternate feature set based on Haar wavelets instead of the usual image intensities. Viola and Jones [Viola and Jones, 2001] adapted the idea of using Haar wavelets and developed the so called Haar-like features. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in these regions and calculates the difference between them. This difference is then used to categorize subsections of an image. For example, let us say we have an image database with human faces. It is a common observation that among all faces the region of the eyes is darker than the region of the cheeks. Therefore a common Haar feature for face detection is a set of two adjacent rectangles that lie above the eye and the cheek region. The position of these rectangles is defined relative to a detection window that acts like a bounding box to the target object (the face in this case).

In the detection phase of the Viola-Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such a Haar-like feature is only a weak learner or classifier (its detection quality is slightly better than random guessing) a large number of Haar-like features are necessary to describe an object with sufficient accuracy. In the Viola-Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier.

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time (approximately 60 microprocessor instructions for a 2-rectangle feature).

Several researchers have studied the impact of in plane rotations for image sensors with the use of twisted Haar-like feature 45^0 [Mita, 2005] or diagonal features fairly good performance has been achieved. These techniques will have little benefit for problems that are sensitive torotations, such as hand identification [Shan, 2004] which are not aligned to fixed angles (0^0, 45^0, 90^0, etc).

Haar-like feature based classifiers like the Jones and Viola, face detector work in almost real time using the integral image (or summed area table) data structure that allows features to be calculated at any scale with only 8 operations. The integral image at location x, y contains the sum of the pixels above and to the left of x, y, inclusive:

$$II(i,j) = \sum_{x' \le x,\ y' \le y} I(x', y'), \quad (1)$$

where $II(i,j)$ is the integral image and $I(i,j)$ is the input (original) image (see Fig. 2 (a)). Using the following pair of recurrences the integral image can be computed in one pass over the original image:

$$\begin{aligned} S(x,y) &= S(x,y-1) + I(x,y), \\ II(x,y) &= II(x-1,y) + S(x,y), \end{aligned} \quad (2)$$

where $S(x,-1) = 0, II(-1,y) = 0.$

Using the integral image any rectangular sum can be computed in four array references (see, Fig.2b). Clearly the difference between two rectangular sums can be computed in eight references. Since the two-rectangle features defined above involve adjacent rectangular sums they can be computed in six array references, eight in the case of the three-rectangle features, and nine for four-rectangle features. The sum of the pixels within rectangle *D* can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle *A*. The value at location 2 is *A* + *B*, at location 3 is *A* + *C*, and at location 4 is *A* + *B* + *C* + *D*. The sum within *D* can be computed as 4 + 1 - (2 + 3).

However standard Haar-like features are strongly aligned to the vertical/ horizontal or 45° (see, Fig. 3.) and so are most suited to classifying objects that are strongly aligned as well, such as faces , buildings etc.

Standard Haar-like features consist of a class of local features that are calculated by subtracting the sum of a sub region of the feature from the sum of the remaining region of the feature.

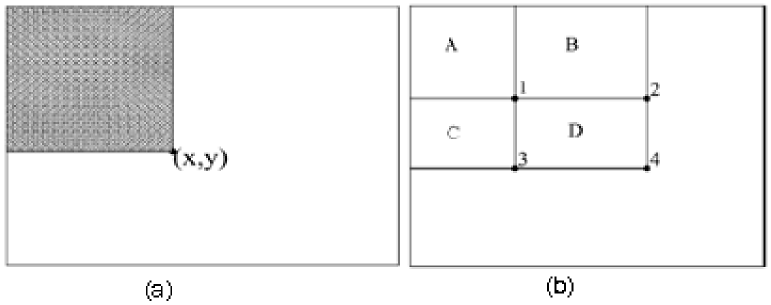

Figure 2. (a) Sum of the integral image at point (*x, y*) and (b) integral image array references

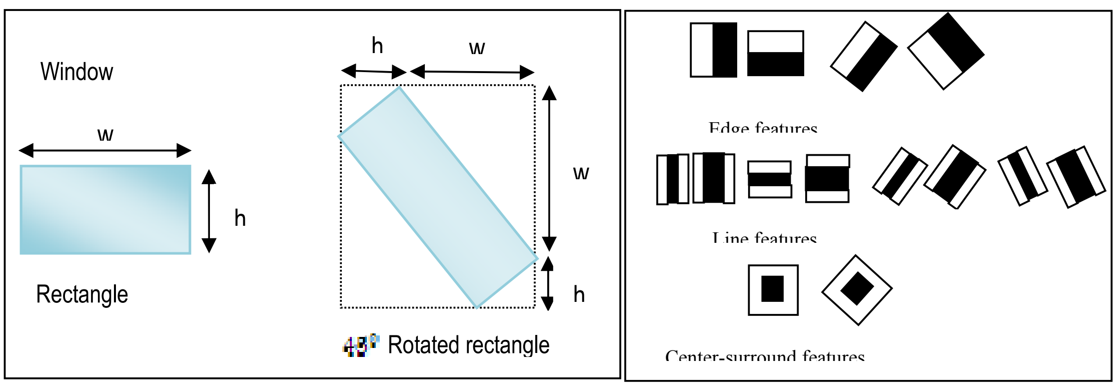

Figure 3. Upright and 45° rotated rectangle, features prototypes of simple Haar-like

Decreasing volume of face images database using LEM

Edge information is a useful object representation feature that is insensitive to illumination changes to certain extent. If the edge detection step is successful, the subsequent task of interpreting the information contents in the original image may therefore be substantially simplified. However, it is not always possible to obtain such ideal edges from real life images of moderate complexity.

In this section we are using $n \times n$ rectangular mask for getting LEM of image (Fig. 3). We are dividing all image into $n \times n$ size of rectangles (in this work n=8) and these rectangles must be smaller than the main features of the image details (nose, mouth, eyes in this case). It helps us to observe all brightness points and avoid from not important pixels of image.

We are using equation (3) for calculating brightness of horizontal or vertical pixels and their opposite side pixels (see, Fig. 4).

$$V = \sum_{i=1}^{8} \left[f(i,1) - f(i,8) \right], \quad H = \sum_{j=1}^{8} \left[f(1,j) - f(8,j) \right] \qquad (3)$$

We could develop function having brightness values of horizontal and vertical sides on current rectangle, which is drawing lines according to brightness values (see, Fig.5).

In order to explore the feasibility of using image LEM in face detection and recognition system we reconstructed all image database (as we mentioned in introduction quantity of images is 40) and get their LEMs.

After that we used OpenCV 2.1 version (Open Source Computer Vision) for detecting and recognizing human faces. We used Haar classifier face detection algorithms functions in which we used 8 futures prototype in Fig. 3. In Fig. 6 given reconstructed images data flow for Haar classifier face detection and recognition algorithm.

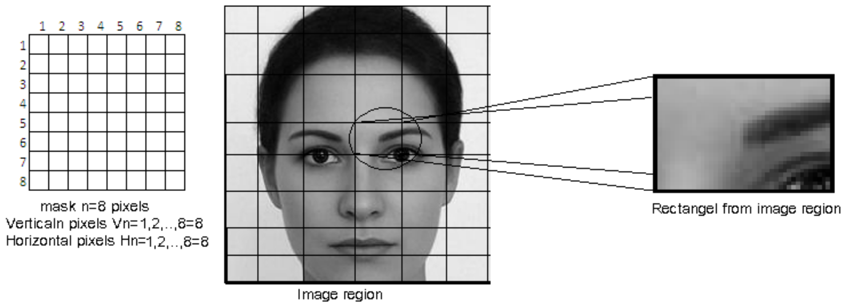

Figure 4. 8x8 mask and image divided with 8x8 rectangles.

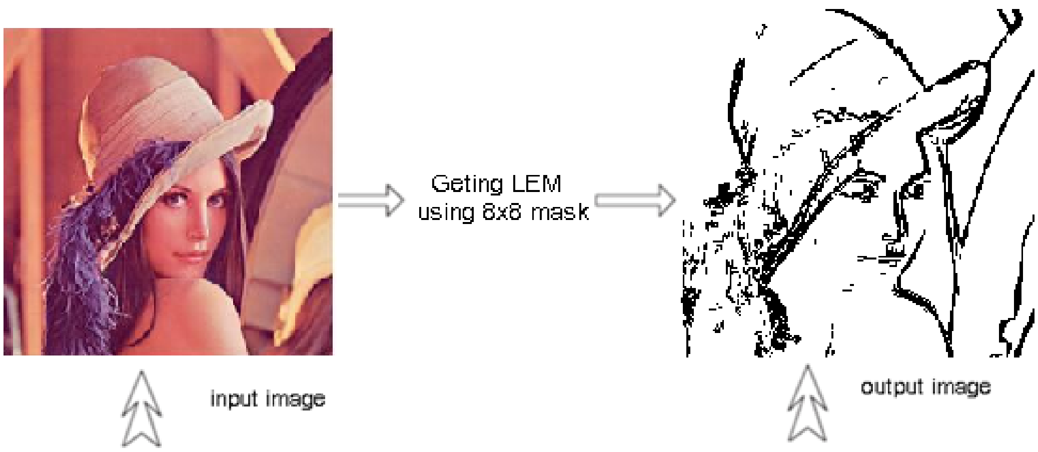

Figure 5. Getting LEM of image using 8x8 mask

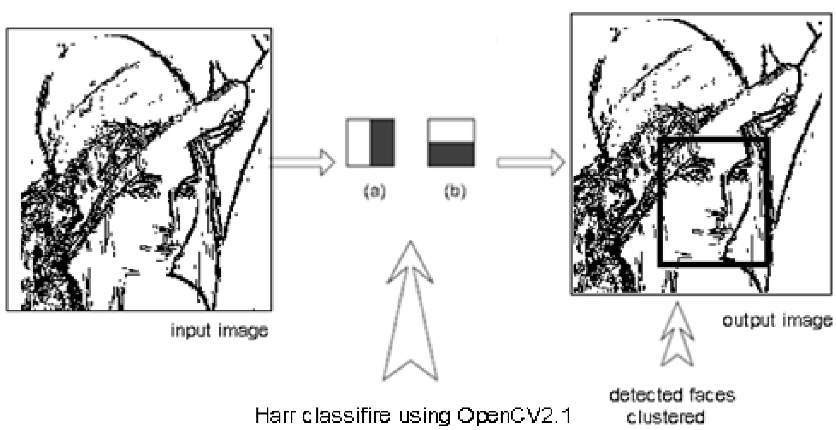

Figure 6. Reconstructed images data flow for Haar classifier

We used Haar classifier face detection algorithms and developed Face detection and recognition software tolls (FDRT) which is using our images LEM database. Comparison results with OpenCV given in table bellow, where we used 40 images of 250×250 pixel sizes.

Training database type	Method/ Tools	Detection and recognition time	Training database size
Color (RGB)	OpenCV2.1	2.02msec	2Mb
Image LEM	OpenCV2.1	2.02msec	1.2Mb
Image LEM	Proposed	1.01msec	1.2Mb

From this table we can say that training database volume reduced ~40 % and time decreased for face detection and recognition. So we get image optimization and face recognition tools which can help us to optimize and reduce the volume of the face images database and recognition time.

Conclusion and future work

In a future work we will try to increase percentage of correctly recognition for real time frame moving face discovering and identifying automation system, which will use result of this work and will try identifying and recognizing faces using database with LEM of images.

Bibliography

[Tolba, 2006] A. S. Tolba, A.H. El-Baz, and A.A. El-Harby Face Recognition: A Literature Review International Journal of Signal Processing, 2, 2006

[Face Database] Face Database, http://vis-www.cs.umass.edu/lfw/

[Takacs, 1998] B. Takacs, "Comparing face images using the modified hausdorff distance," Pattern Recognition, vol. 31, pp. 1873-1881, 1998.

[Gao, 2002] Y. Gao and K.H. Leung, "Face recognition using line edge map," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, No. 6, 2002.

[Leung, 1990] M.K.H. Leung and Y.H. Yang, "Dynamic two-strip algorithm in curve fitting," Pattern Recognition, vol. 23, pp. 69-79, 1990.

[Bern, 2002] Bern Univ. Face Database, ftp://iamftp.unibe.ch/pub/Images/FaceImages/, 2002.

[Purdue Univ. Face Database] Purdue Univ. Face Database, http://rvl1.ecn.purdue.edu/~aleix/
aleix_face_DB.html, 2002.

[Konstantinos, 2008] Konstantinos Koutroumbas and Sergios Theodoridis. Pattern Recognition (4th ed.). Boston: Academic Press. 2008

[Duda, 2001] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. Wiley, 2001

[Shan, 2004] C. Shan, Y. Wei, T. Tan, F.Ojardias , "Real time hand tracking by combining particle filtering and mean shift", Proc. Sixth IEEE Automatic Face and Gesture Recognition, FG04, pp: 229-674, 2004.

[Papageorgiou, 1998] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection", International Conference on Computer Vision, 1998.

[Viola and Jones, 2001] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. 511-518, 2001.

[Mita, 2005] T. Mita, T. Kaneko, and O. Hori, "Joint Haar-like features for face detection," Tenth IEEE International Conference on Computer Vision: ICCV 2005. Vol. 2, pp. 1619-1626, 2005.

[Yang, 1996] Jie Yang , A. Waibel, "A real-time face tracker", Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96), p.142, Dec., 1996

[Raul, 1996] Raul Rojas, Neural Networks - A Systematic Introduction, Springer-Verlag, Berlin, New-York, 1996, 502p.

Authors' Information

Grigor Poghosyan *– PHD student, Institute for Informatics and Automation Problems of NAS of RA; e-mail: grigor.poghosyan@mail.ru*
Major Fields of Scientific Research: Digital Signal and Image Processing

Hakob G. Sarukhanyan *– Head of Digital Signal and Image Processing Laboratory, Institute for Informatics and Automation Problems of NAS of RA; e-mail: hakop@ipia.sci.am*
Major Fields of Scientific Research: Digital Signal and Image Processing, Fast Orthogonal Transforms, Parallel Computing