COMPRESSION AND TRANSMISSION OF FACIAL IMAGES OVER VERY NARROWBAND WIRELESS CHANNELS

Aysegul Gunduz, Hamid Krim, and P. Allan Sadowski[†]

North Carolina State University, Raleigh, NC 27695-7914
[†]North Carolina State Highway Patrol, Raleigh, NC 27699-4702
{agunduz,ahk}@eos.ncsu.edu, asadowski@ncshp.org

ABSTRACT

Law enforcement officers on mobile duty are often confronted with ID authentication of subjects entailing the transmission of a driver's license picture over wireless channels that are very narrowband. To access mug shots in a reliable and timely manner, a real time compression and decompression method with high compression ratios is required at the server database and the mobile client unit. The technique presented in this paper minimizes the size of the data sent over the channel by locally storing common features of the human face in the client computers. Pre-processing of server database images, such as facial feature extraction, are used to extract these common facial features, obtained via ravines and image singularities. The implemented file transfer protocols are based on basic TCP/IP client-server models and make use of socket programming. Experimental results show a 5x improvement in transfer time over typically saturated channels.

1. INTRODUCTION

Accessing mug shots of subjects in a timely and reliable way is critical to law enforcement agencies, especially to the officers on mobile duty. A mug shot picture from databases (e.g. Dept. of Motor Vehicles) in hand provides officers with person identification in cases such as absence of an ID or authentication in case of a suspicion. Wireless transmission is the solution for communication with mobile computers. However, the wireless network channels of law enforcements have to be shared among various agencies and the channels run almost always at full capacity. Standard compression algorithms (e.g. JPEG 2000) [8] exhibit an unacceptable performance in so far as the required transmission time is concerned (upload of 4 to 5 min). We have thus had to develop a highly specialized technique optimized around our mug shot wireless transmission over this highly constrained network. The objective of this paper is to describe a robust, efficient, and simple image compression technique optimized for ID pictures and using a communication protocol used by the North Carolina State Highway Patrol over a very narrowband channel (<10 kbps).

Image compression addresses the problem of reducing the amount of data to represent an image. Yet, in the case of transmission, the goal is to reduce the amount of data to be transmitted. If a common portion of the data of all images were to be permanently available at the destination, then the differentiating data among pictures need be transmitted. That is precisely the basis of our approach as we describe next.

Human faces share a common template, which is composed of two eyes, one nose, one mouth, and so forth. Moreover, their locations relative to each other on the average are known. Only shape, size and distance between these features vary from person to person. The methodology presented in this paper is based on using common features of face images, i.e. finding a template at some level that is common to all faces that will be available at all client computers.

This paper is organized as follows. Section 2 gives a brief description of the wireless network and its schematic. Section 3 describes the construction of a common image template from a given database and details the feature extraction technique. The compression results and the attained transmission statistics are provided in Section 5. Finally, concluding remarks are presented in Section 6.

2. THE WIRELESS NETWORK CONSTRAINTS

The wireless network infrastructure of North Carolina Criminal Justice Information Network (CJIN) is managed and operated by the State Highway Patrol (NCSHP). This network serves to send information to mobile trooper and police cars, which are equipped with portable computers and radio modems. With over 275 agencies, there are almost 8000 officers registered to the CJIN wireless network. During peak usage, the approximate number of users in the network is 5000, with an average of 2500 daily/evening users - dropping to 2000 during early morning hours.

The radio modem in the mobile vehicle, communicates with the nearest base tower at a low rate 19.2kbps shared channel. Actual throughput on the channel is approximately 10kbps - the actual rate varying due to range and S/N issues. The radio tower is connected to the state network via a 14.4kbps/56K line through radio network controllers and radio switches. The overall transmission schematic is shown in Figure 1.

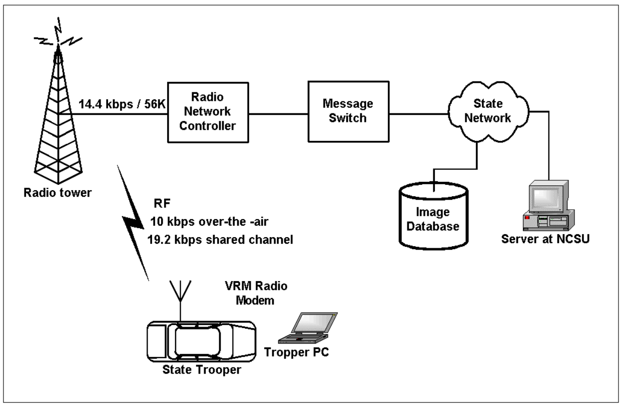

Figure 1. The end-to-end image transmission schematic.

Because of the very narrow bandwidth of the network, the heavy traffic flow, and the high demand of image transmission, any algorithm would be subject to the existing infrastructure, and the resulting compression and decompression techniques should be highly efficient and operating at real time. Note that, since the goal is facial recognition of a person by an officer, lossy compression is acceptable as long as identifiability is preserved.

The file transfer protocol is based on the TCP/IP client-server model and is implemented using socket programming. The server is programmed for multiple clients and can handle up to 2000 users.

3. FACIAL IMAGE COMPRESSION

The color images in the database are first converted to grayscale. The images originally contain redundant regions with no information about the face that can be cropped. Automatic cropping of a large database of pictures is developed. Figure 2(a) shows a database image that is converted to grayscale.

3.1. Finding the mean of the database

It may be observed from the database that some images or people have more in common than others. People with glasses or mustaches, Caucasians etc. are such examples. The database therefore, is divided into groups depicting these subclasses. The mean of each subclass is obtained in order to extract the common, coarse template of the group. The challenge in obtaining this template denoted by C(x,y) lies in the variability of pictures as a function of distance, illumination and pose at C. A resizing and alignment algorithm was developed on the basis of a triangle with the mouth and the two eyes as vertices.

3.2. Feature Extraction

The two dimensional image I(x,y) may be considered as a surface, S = { (x,y,z) : z = I(x,y) } , where the values of the z-axis is represented by the image intensity at the pixel (x,y). The feature extraction approach is based on extracting ravines and image singularities of the surface S [3], [4]. Ravines are points on the surface a local maximum in the corresponding principal direction [3], [6]. To reduce the effect of noise while preserving the singularities of interest, an anisotropic diffusion is applied to each image [7].

The most important and characterizing curvatures of a surface are the *Gaussian* and *mean curvatures* that can be formulated respectively as follows [4],[6]

$$K = \det(F_1^{-1}F_2)\,, \tag{1}$$

and

$$H = \frac{1}{2}Tr(F_1^{-1}F_2), \tag{2}$$

where $F_1^{-1}F_2$ is the matrix of the shape operator, and F_1 and F_2 are the first and second fundamental forms given respectively by [4],[6]

$$F_1 = \begin{bmatrix} 1+I_x^2 & I_xI_y \\ I_xI_y & 1+I_y^2 \end{bmatrix}, \tag{3}$$

and

$$F_2 = \frac{1}{\sqrt{1+|\nabla I|^2}}\begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix}, \tag{4}$$

where the partial derivatives, I_x, I_y and I_{xy} are attained using recursively implemented Gaussian kernels [2].

Another important curvature and the one our approach is concerned with is the *maximum curvature*, which is defined as

$$k_{\max} = \max\left(H + \sqrt{H^2 - K}\right) \tag{5}$$

and the corresponding direction $\vec{d}_{\max}$ is called the *direction of the maximum curvature.*

The *directional derivative of the maximum curvature* (DMC) is given by [6]

$$DMC = \nabla k_{\max} \cdot \vec{d}_{\max}. \tag{6}$$

Given that the image singularities at the pixels {(x,y):DMC=0}, the ravines may be represented with the following [6]:

$$R(x,y) = \begin{cases} 1, \; if \; DMC(x,y)=0, and \;\; k_{\max}(x,y) > 0 \\ 0, \; otherwise \end{cases}. \tag{7}$$

To locate the eyes, a box of the facial region is extracted from the database images. The extraction of the box is automatic and the eyes are always contained within the box. Figure 2(b) shows the box extracted from a database image. The ravines of the facial region detect the eyes since; the eyes are local extremes (Figure 2(d)). This figure is further processed to extract the eyes by first finding possible eye pairs, and then verifying the pair based on the geometry and relative position (Figure 2(e)).

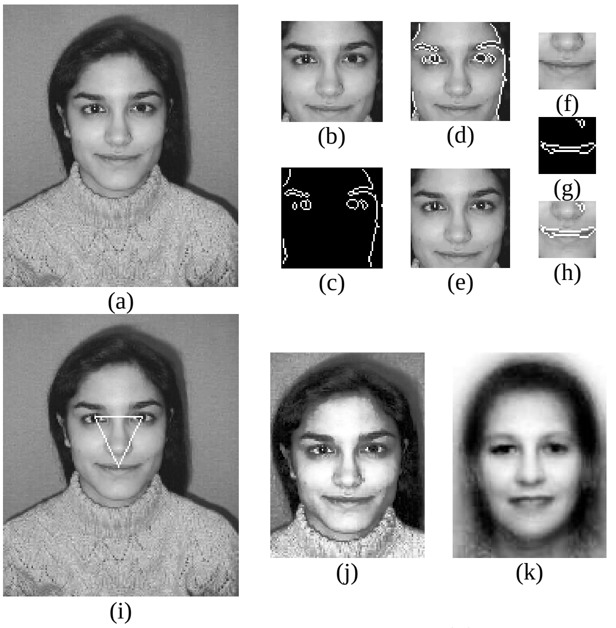

Figure 2. (a) Original image**, (b)** Extracted box of facial region**, (c)** Ravines, R(x,y)**, (d)** Superposition of the original image and ravines**, (e)** The center of gravity of the extracted eye regions are marked with white pixels**, (f)** Extracted box of mouth region**, (g)** Ravines of the mouth**, (h)** Superposition of the mouth and its ravines**, (i)** The triangle connecting the located features**, (j)** The resized, cropped, aligned, and histogram equalized output image**, (k)** The average image of a group of 100 Caucasians.

Once the eyes are located, using the row information of the eyes we are able to detect the mouth as shown in Figure 2(f). For space sake, we defer the details to [3]. Figure 2(h) shows the located eyes and mouth and the triangle connecting them.

The optimal triangle sizes are obtained through experimentation over a large database. The illumination variability is greatly reduced by histogram equalization of the aligned and cropped images. Figure 2(i) illustrates such an output image. Figure 2(j) shows the average image of a group of 100 Caucasians.

3.3. Image representation by level sets

Level sets are the "binary shadows" of an image and can be defined by

$$L_\lambda^{(I)}(x,y) = \begin{cases} 1, \ if \ I(x,y) \geq \lambda \\ 0, \ otherwise \end{cases} \tag{8}$$

where I(x,y) is the image of interest. The family of level sets is defined for all values of λ in the range of I and provides a complete, contrast invariant representation of the image [5]. Hence, an image may be decomposed into its level sets and recovered.

The *principal level sets* are the level sets that contain the basic information required to recover the image. We have shown that the facial images in the NCSHP database can be represented by 6 principal level sets, $L_\lambda^{(I)}$, $\lambda \in$ {90, 120, 150} and $L_\lambda^{(C)}$, $\lambda \in$ {30, 60, 180}, where I and C range from 0 to 255. The common, coarse level sets extracted from the mean image, $\{L_{30}^{(C)}, L_{60}^{(C)}, L_{180}^{(C)}\}$, (i.e. the common template) can be saved in the client computers and need not to be sent over the wireless channel.

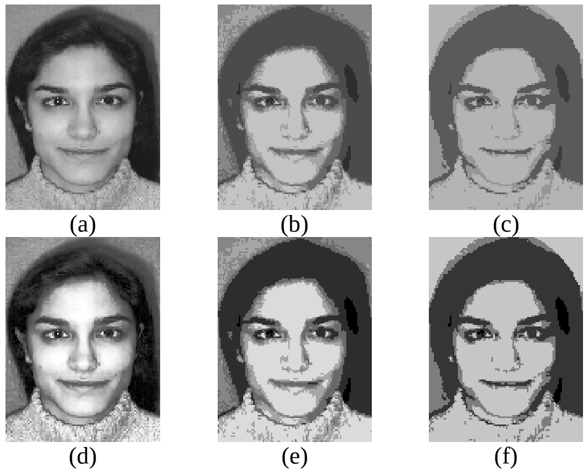

Figure 3.(a) Crop of the original image**, (b)** Image recovered from its own 6 principal level sets**,(c)** Image recovered using 3 principal level sets of the mean image**, (d)** Resized, cropped, aligned, and histogram equalized image **, (e)** Image recovered from its own level sets**, (f)** Image recovered using 3 level sets of the mean image.

3.4. Compression of differential level sets

The non-common level sets that are to be transferred, $\{L_{90}^{(I)}, L_{120}^{(I)}, L_{150}^{(I)}\}$, are first 8-bit run-length encoded at the server. This is followed by a real-time Lempel-Ziv-Oberhumer (LZO) Compression, which favors speed over compression ratio.

3.5. Decompression and Recovery

At the client side, the received level sets are decompressed in the reverse order: LZO decompression followed by run-length decoding. The image is recovered from the level sets by

$$\hat{I} = \max\{\lambda : L_\lambda = 1\} \tag{9}$$

where $\lambda \in$ {30, 60, 90,120,150, 180}.

Figures 3(a) and 3(d) have been decomposed into their principal level sets along with the mean image of Figure 2(j). Figures 3(b) and 3(e) show the images recovered from their own levels sets. In Figures 3(c) and 3(f) the images are recovered using three level sets from the mean image.

	Original Size (in KBytes)	Compressed Size (in Bytes)	Compression Ratio	Average Transmission Time of Original Image	Average Transmission Time of Compressed Image	Improvement in Transmission Time
Image 1	7.30	806	88.96%	20.75 sec	4.6 sec	4.51
Image 2	7.90	966	87.77%	23.25 sec	6.0 sec	3.87
Image 3	8.19	685	91.64%	26.75 sec	5.0 sec	5.35
Image 4	8.41	928	88.97%	24.25 sec	4.8 sec	5.05
Image 5	8.66	780	90.99%	25.50 sec	5.8 sec	4.40
Image 6	8.70	841	90.33%	26.25 sec	6.4 sec	4.10
Image 7	9.19	1065	88.41%	27.00 sec	6.0 sec	4.50
Image 8	9.31	827	91.12%	26.50 sec	4.8 sec	5.52
Image 9	9.59	835	91.29%	25.00 sec	4.8 sec	5.21
Image 10	9.67	1053	89.11%	24.50 sec	5.8 sec	4.22
Image 11	9.84	1087	88.95%	35.00 sec	5.8 sec	6.03
Image 12	9.95	928	90.67%	32.00 sec	6.4 sec	5.00
Image 13	10.00	811	91.89%	28.75 sec	5.4 sec	5.32
Image 14	10.50	780	92.57%	32.75 sec	4.4 sec	7.44
Image 15	10.70	791	92.61%	27.25 sec	5.0 sec	5.45
Image 16	10.80	819	92.42%	29.50 sec	4.6 sec	6.41
Image 17	11.30	767	93.21%	32.50 sec	4.8 sec	6.77
Image 18	11.60	1291	88.87%	37.00 sec	6.4 sec	5.78
Image 19	12.10	1052	91.31%	36.25 sec	5.6 sec	6.47
Image 20	12.20	803	93.42%	41.25 sec	4.6 sec	8.97

Table 1. Compression results and transmission time improvements of 20 typical images from the group of 100 white Americans.

4. COMPRESSION RESULTS and TRANSMISSION STATISTICS

The average size of the driver's license images in the database is approximately 10Kbytes. The average compression ratio of the implemented algorithm was found to be 90.73%. Hence, the average size of the file that has to be transferred is reduced to 927 Bytes. The effect of this reduction can be seen in the obtained transmission statistics.

The transmission statistics presented were taken on different time slots of different days. While, the average transmission time of original database images was found to be 29.085sec, the average transmission time of the compressed images was found to be 5.35sec.

5. CONCLUSIONS

Table 1 shows the compression results and the transmission times of 20 images, which is a good set with variety to represent the group of 100 Caucasians. It may be inferred from the table that the compression algorithm is not linear. Yet, the compression ratios have led to a 5x improvement in the transmission times over the wireless network.

With the compression technique described in this paper, the transmission of facial images has been made much faster. The image quality is acceptable as long as the images can help the law enforcements identify criminals.

6. REFERENCES

[1] J. F. Canny, "Finding edges and lines in images", Technical Report, MIT, Artificial Intelligence Laboratory, Cambridge, Massachusetts, June 1983.

[2] R. Deriche, "Recursively Implementing the Gaussian and its derivatives", *Proc. IEEE Int. Conf. on Image Processing*, Singapore, pp. 263-267, September 1992. Also INRIA Research Report 1893, 1993.

[3] A. Gunduz, H. Krim, "Facial Feature Extraction Using Topological Methods", *submitted ICIP 2003.*

[4] A.B. Hamza, H. Krim, "A topological variational model for image singularities", *Proc. IEEE Int. Conf. on Image Processing*, Rochester, NY, vol.1, pp. 369-372, September 2002.

[5] C. Kervrann, M. Hoebeke, and A. Trubuil, "A level line selection approach for object boundary estimation", *Proc. IEEE Int. Conf. on Computer Vision*, Greece, vol.2, pp. 963-968, September 1999.

[6] O. Monga, N. Armande, and P. Montesinos, "Thin nets and crest lines: application to satellite data and medical images", *Computer Vision and Image Understanding*, vol.67, pp. 285-295, September 1997.

[7] P. Perona, J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.12, no.7, pp.629-639, July 1990.

[8] Taubman, D., M.W. Marcellin, *Jpeg2000:Image Compression Fundamentals, Standards, and, Practice,* Kluwer Academic Publishers, Boston, November 2001.