2005

# Region segmentation for facial image compressing

Alexander Tropf
*Edith Cowan University*

Douglas Chai
*Edith Cowan University*

10.1109/ICICS.2005.1689320 This conference paper was originally published as: Tropf, A. S., & Chai, D. K. (2005). Region segmentation for facial image compressing. Proceedings of Fifth International Conference on Information, Communications and Signal Processing. (pp. 1565-1569). Bangkok, Thailand. IEEE. Original article available here

# Region Segmentation for Facial Image Compression

Alexander Tropf and Douglas Chai
Visual Information Processing Research Group
School of Engineering and Mathematics, Edith Cowan University
Perth, Australia

*Abstract*— **This paper addresses the segmentation of passport images in order to improve quality of significant regions and to further reduce redundancy of insignificant ones. The approach is to first segment a facial image into two major regions, namely background and foreground. Here a new technique using pixel difference is presented. To compress facial regions at better quality, a face segmentation algorithm is introduced that detects eyes and mouth in a face. Region of Interest (ROI) coding is then used to obtain better quality for facial features. At the end, some strategies that make use of region segmentation are proposed in order to increase performance in entropy coding.**

## I. Introduction

The increasing terrorist threat over the last few years has created a greater demand for new and efficient automatic surveillance systems in public facilities. One possibility of such a surveillance system is a face recognition system, where the face of a person is detected, analysed and compared to an image stored in a database. Depending on task and number of images, such databases can be extremely large. A digital colour image of size $128 \times 100$ pixels at 24 bit/pixel requires a storage size of about 38 KB in raw format. Todays standard compression algorithms are able to greatly reduce storage of such a facial image, however since there is no difference made between insignificant regions like background and significant ones like important facial features, compression efficiency remains low. This paper presents methods to segment a passport image into different regions which allows the coder to compress background at a higher rate and facial regions with sufficient quality for the face recognition system.

To show the effect of the proposed region segmentation approach for facial image compression, Vector Quantisation (VQ) [1], [2] is used with the Enhanced Linde-Buzo-Gray (ELBG) algorithm [3].

This paper is organised as follows. Background separation is described in Section II. Section III presents the face segmentation algorithm. Experimental results are presented in Section IV, while conclusion is given in Section V.

## II. Background Separation

Typical passport images have the following characteristics and requirements[1]:

1) The complete head of a person including throat and top of the shoulders are visible.

[1]According to Australian Passport Photo Guidelines (http://www.passports.gov.au/Web/requirements/photos.aspx), Aug 2005 and UK Passport Service (http://www.passport.gov.uk/downloads/PLE03_web_Mar05.pdf), Mar 2005.

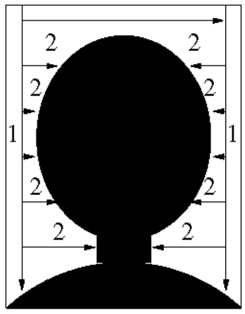


Fig. 1. Illustration of background scan: step (1), search lowest background pixel above shoulder; step (2), scan every row from top to start of shoulder

2) The picture size is of 45 mm (H) $\times$ 35 mm (W), and the face takes 65-80% of the picture height.
3) The background is uniform, light coloured and gives good contrast to facial features.
4) The person has to look straight into the camera with a neutral facial expression.

Passport images are in colour, however, the proposed algorithms require only the luminance component. Therefore an RGB image is transformed into YCbCr colour space and only the Y-component is used.

To detect the background region in the luminance component of a passport photo one can use the fact that it has to be uniform. Two neighbouring image samples $s_i$ and $s_{i+1}$ can be considered as uniform if their absolute difference $\left|\Delta d_{s_i,s_{i+1}}\right|$ is below a background threshold $\epsilon_{bg}$. To reduce complexity in the 2-D spatial image plane to a 1-D problem, only single rows or columns of image samples or pixel shall be considered. According to the spatial characteristics of passport images, the first and last column contains background from top of the image to the shoulder of the person. The last background sample just above the shoulder can be detected by scanning the first and last column until $\left|\Delta d_{s_i,s_{i+1}}\right| \geq \epsilon_{bg}$. Next, for each row that starts with a background pixel, a row scan is performed from the left image border to the right, calculating differences of pixel pairs. The edge is reached per definition, whenever $\left|\Delta d_{s_i,s_{i+1}}\right| \geq \epsilon_{bg}$. Then all previous pixels in that row are classified as background. If not the complete row consists of background, one proceeds by scanning from the right image border to the left until $\left|\Delta d_{s_i,s_{i+1}}\right| \geq \epsilon_{bg}$. Again, from the stop point on to the right image border, all pixels are classified as background pixel (see Figure 1).

The background separation algorithm is fast and simple to realise as differences are easy to compute with hardware. However, the algorithm is not very robust towards small variations of $\epsilon_{bg}$ and can fail for rows that have smooth edges between background and foreground region. This is because only one neighbouring pixel is taken for comparison with a current pixel. Hence one can say a memory of $m = 1$ is used to store the previous pixel. By increasing that memory ($m > 1$) and taking more previous pixels into account, the algorithm should improve. To give the previous samples $s_{i-m}$, $\dots$,$s_{i-1}$ different weights, the weight factors $w_{i-m}$,$\dots$,$w_{i-1}$ are introduced. An average previous sample $\bar{s}_{i-1}$ is calculated as

$$\bar{s}_{i-1} = \sum_{k=1}^{m} \frac{s_{i-k} \cdot w_{i-k}}{\sum_{k=1}^{m} w_{i-k}} = \sum_{k=1}^{m} s_{i-k} \cdot w_{i-k}^{*} \tag{1}$$

where $w_{i-k}^{*}$ refers to the normalised weights. By looking at a typical passport image, one can say that the closer a sample is to the right or left image border, the higher is the probability that this sample belongs to background. Therefore an exponential weight function $w_{i-k} = 2^k$, $k = [m, \dots, 1]$ is used, which gives an image sample, that is furthest away from the current sample, the maximum weight.

Finally, a binary separation mask is obtained where "zero" indicates background and "one" foreground.

## III. Face Segmentation

Once the background has been identified, its redundancy can be reduced to a minimum. In VQ this is done by representing background with only one code vector. The remaining foreground region is consisted of shoulder and the head of the person in the image. Facial features like eyes and mouth are extremely important for the face recognition system and require a higher quality, whereas other parts of the foreground region, like shoulder and hair are less important and could be compressed stronger. Therefore the next task is to define an area around eyes and mouth, it will be referred to as the Primary Region of Interest (PROI), which is later encoded with high quality. To extract the PROI of a facial image, it is necessary to detect the eyes and mouth. Most existing eye and mouth localisation methods are template based [4], [5]. However a simpler eye mouth detection algorithm shall be proposed here, which will utilise the fact that the position of the face is approximately known after background separation.

The pupil of a human eye generally has a low luminance intensity. Most human skin colour that surrounds the eye has luminance intensity that is higher than the one of the eye. Hence the eye causes a local minimum in the image plane that should be possible to detect. The same is valid for the mouth, as long as the person has a neutral facial expression, which is one of the passport image requirements as mentioned in Section II.

To simplify the search for eyes and mouth, the initial search area should be as small as possible. Here the background separation algorithm, as discussed in Section II, will help to eliminate background and give the top, left and right limit of the head. For the bottom point a fixed threshold is defined that is below the mouth for all images. Since spatial features in passport images are very similar, one can further tighten the limits to obtain a "coarse facial search area" which contains only face and, depending on the person, hair features.

Next, a vertical profile of the luminance data is calculated by taking the mean of every row in the search area. A horizontal profile can be obtained by computing the average of the columns. To classify the right minima, an elimination procedure shall be introduced that discards insignificant minima in several steps. First, the expression for local maximum and minimum in an image profile needs to be defined. Let $p(i)$ and $p(j)$ be the discrete profile functions of a vertical and horizontal profile, respectively. According to [6], for a set of local maxima, $P_{max}$, in any profile $p(k)$ with a length $a$, the condition

$$\begin{aligned} P_{max} \;&=\; \{(k, p(k)) \mid p(k) > p(k-1) \wedge \\ &\quad\; p(k) > p(k+1), 2 \le k \le a-1\} \end{aligned} \tag{2}$$

has to hold true. A set of local minima, $P_{min}$, is given by

$$\begin{aligned} P_{min} \;&=\; \{(k, p(k)) \mid p(k) < p(k-1) \wedge \\ &\quad\; p(k) < p(k+1), 2 \le k \le a-1\}. \end{aligned} \tag{3}$$

Research has shown that there exist points $p(k)$ that have a neighbour that holds the condition $p(k+1) = p(k)$. If $p(k) < p(k-1) \wedge p(k) < p(k+2)$ is true, then $p(k)$ should still be defined as minimum. For such a case, (3) can be rewritten as

$$\begin{aligned} P_{min} \;&=\; \{(k, p(k)) \mid p(k) < p(k-1) \wedge \\ &\quad\; p(k) < p(k+1) \vee \\ &\quad\; p(k) < p(k-1) \wedge \\ &\quad\; p(k) = p(k+1) \wedge \\ &\quad\; p(k) < p(k+2), 2 \le k \le a-2\}. \end{aligned} \tag{4}$$

Since a person's eyes will cause low minima, one can discard high local minima that occur close to the left side of the vertical profile (corresponds to top part in search area). The vertical profile can be further cut, by discarding high local maxima that appear close to the top of the search area. That way the vertical profile is shortened enough that the next minima from the left either belongs to eyebrows or eyes.

The mouth of a facial image appears as a very strong local minimum in the vertical profile. However the profile can be distorted by dark hair structures from the left and right side. Hence it is better to have a very narrow horizontal search area around the mouth to calculate the vertical profile with. To obtain this area it is assumed that the horizontal centre of the mouth is close to the horizontal centre of the nose. The nose can be easily identified in the horizontal profile as the closest maximum of the profile centre. Next, the area $\pm\beta$ around the centre point defines the horizontal width of the search area that is used to calculate the new vertical profile. Of course $\beta$ has to be chosen in a way that the mouth is included for all images of a database. Then all minima on the right side of the vertical profile (lower half in search area) are considered as "mouth candidate" (see Figure 2). Tests have shown that the strongest

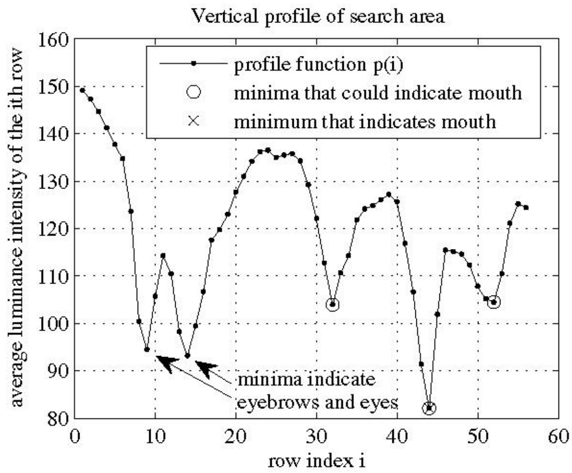


Fig. 2. Vertical profile for mouth and eye detection

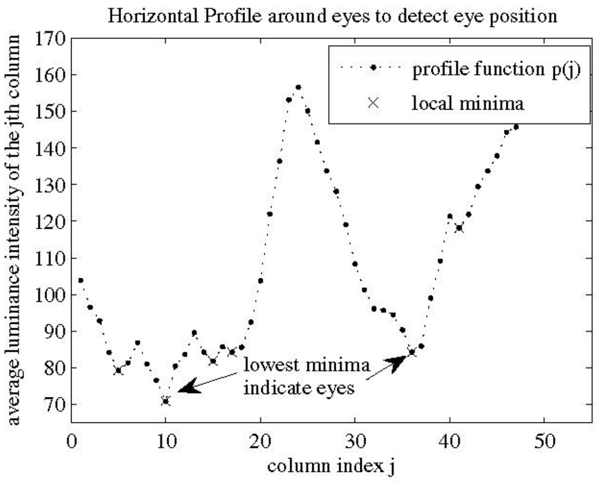


Fig. 3. Horizontal profile for eye detection

minimum always indicates the vertical position of the mouth. The "strongest minimum" is defined as the minimum that has the highest slope between two previous and two consecutive values. Mathematically the sum $s$ of those absolute slopes is

$$\begin{aligned} s \;=\; & |p(i)-p(i-1)|+|p(i)-p(i-2)|+ \\ & |p(i)-p(i+1)|+|p(i)-p(i+2)|. \end{aligned} \tag{5}$$

The minimum with the highest $s$, is then chosen to represent the vertical position of the mouth (see Figure 2). Figure 2 shows two distinctive minima on the left side which correspond to eyebrows and eyes. However not all passport images show a minimum for the eyebrows (e.g. people with blond hair). In such a case there will be only one distinctive minimum present on the left half of the vertical profile, presenting the eye. Knowing these features, one can eliminate less distinctive minima through evaluation of (5) and obtain the vertical position of the eyes. To obtain more distinctive minima of the eyes in the horizontal profile, the mean of only seven rows around the vertical eye-position is taken. The lowest minima in the left and right part of this profile (see Figure 3) corresponds to the horizontal position of the left and right eye, respectively.

After the position of eyes and mouth is known, one can choose the PROI such that those features are just included. Region of interest coding of the PROI will then increase quality around facial features.

## IV. Results

The proposed background separation algorithm has a success rate of 100% as long as the passport image characteristics as stated in Section II are adopted in a strict sense. The simple algorithm ($m = 1$) successfully separates foreground and background of passport images for $\epsilon_{bg} = 10$. Using $m = 4$ and the exponential weight function from Section II makes the algorithm slightly slower, but highly increases its robustness by allowing $\epsilon_{bg}$ to be in the range of 10 to 25. To compress

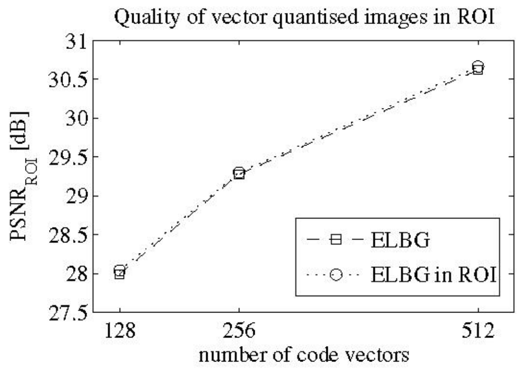


Fig. 4. Quality of vector quantised images with and without background separation

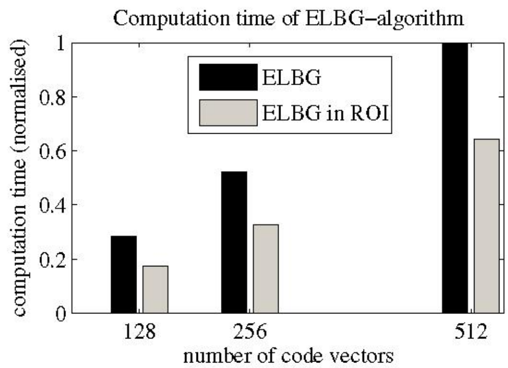


Fig. 5. Computation time of ELBG-algorithm with and without background separation

background at the highest possible rate with VQ, only one code vector is used for its representation. Then the ELBG-algorithm is only applied to code vectors in the foreground region (ROI). The result is a small increase in foreground quality ($PSNR_{ROI}$) as shown in Figure 4. On the other hand, computation time of the ELBG-algorithm decreases by about 35% (see Figure 5).

The parameters of the face segmentation algorithm as proposed in Section III were determined manually through experiments. The algorithm successfully detects eyes and mouth for most images of the image database that corresponds to the passport photo guidelines. Problems in the mouth detection occurred when the person has a beard that causes a minimum below the mouth. However in such a case the error can be limited, by using the vertical eye position and the fact that the distance between eyes and mouth is similar. The segmentation of the vertical eye position was successful in all cases of the database. The horizontal eye detection was successful as long as the passport was taken under uniform illumination (no shades around eyes). The algorithm fails in case a person has very dark skin colour. However even in that case, mouth and eyes appear as distinctive minima. Therefore, by adjusting the parameters, eye and mouth detection can work. The ELBG-algorithm with a weighted MSE [2] was used to achieve higher quality coding in facial regions. A reasonable weight factor is in the range of $2 \ldots 10$ which leads to an average increase in $PSNR_{PROI}$ of 0.5 dB ... 2 dB (see Figure 6a) and a decrease in the overall ROI region of 0 dB ... 0.5 dB (see Figure 6b).

The fact that background is presented by only one code vector can be utilised in entropy coding of the vector quantised images. Figure 7 shows two possible approaches that increase performance of standard Huffman Coding [7]. The first method uses Run-Length coding [8] of the background vector indices. The second one even further reduces redundancy by encoding the differences of the start position of the ROI in each row.

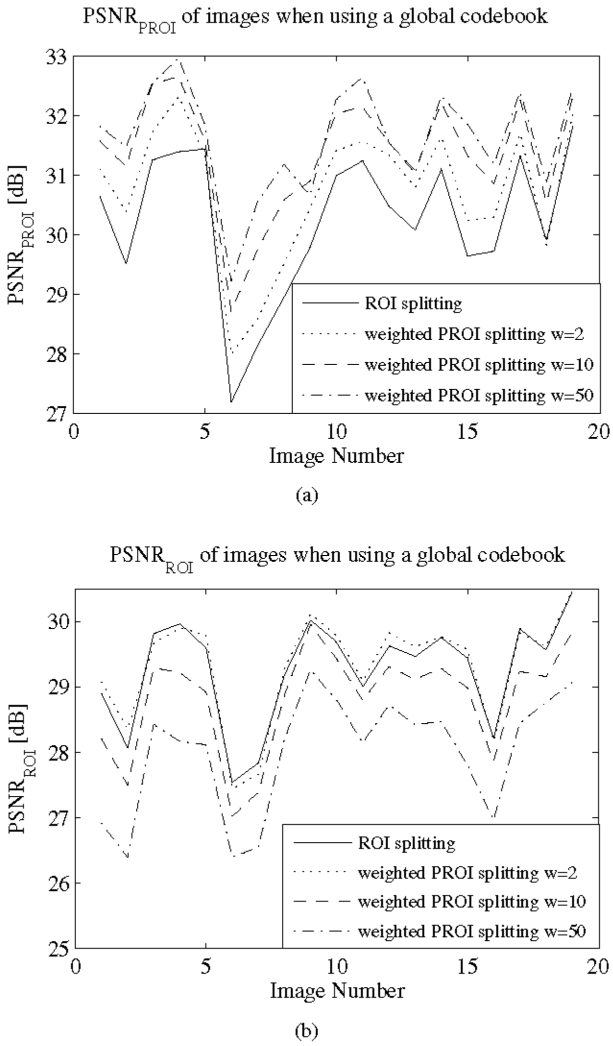


Fig. 6. Results of better quality coding in face region

## V. Conclusion

The proposed background separation algorithm is easy to implement and fast, since the most frequently used operation is a subtraction of numbers. The application allows to decrease background redundancy, while encoding quality of foreground can be increased. By using the compression algorithm only on the foreground region, encoding time can be significantly decreased.

Also, the proposed face segmentation algorithm is fast and easy to implement. The future work is to extent the algorithm from luminance to all colour components to increase robustness. The resulting ROI coding in facial regions leads to a significant increase of quality for facial features.

Finally it could be shown that region segmentation can improve entropy coding of the quantised image data. The outcome is that redundancy of the less important background region can be reduced to an absolute minimum.

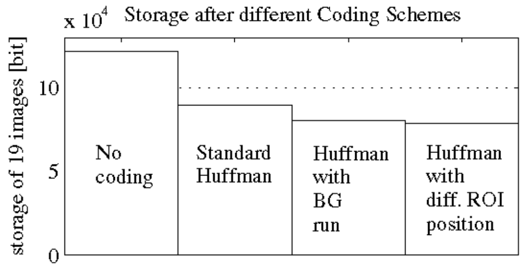


Fig. 7. Results of region segmentation for entropy coding

## REFERENCES


[1] J. Huang and Y. Wang, *Compression of Color Facial Images using Feature Correction Two-stage Vector Quantization*, IEEE Trans. Image Processing, Vol. 8, No. 1, pp. 102 - 109, Jan. 1999.

[2] Majid Rabbani, Paul W. Jones: *Digital Image Compression Techniques,* Tutorial Texts in Optical Engineering, Volume TT 7, 1991.

[3] G. Patanï£¡and M. Russo, *The Enhanced LBG Algorithm*, Neural Networks, vol. 14 no. 9, pp 1219–1237, November 2001.

[4] W. Huang, Q. Sun, C.-P. Lam, and J.-K. Wu, *A Robust Approach to Face and Eyes Detection from Images with Cluttered Background*, ICPR, vol. 1 , pp. 110-114, Aug. 1998.

[5] F. Smeraldi, O. Carmona, and J. Bigun, *Saccadic search with Gabor features applied to eye detection and real-time head tracking,* Image and Vision Computing, vol. 18, no. 4, pp. 323-329, 2000.

[6] Heng-Da Cheng and Ying Sun, *A Hierarchical Approach to Color Image Segmentation Using Homogeneity,* IEEE Trans. Image Processing, Vol.9, No. 12, December 2000.

[7] D. Salomon, *Data Compression: The Complete Reference,* 2nd Edition, Springer-Verlag New York, 2000.

[8] William B. Pennebaker, Joan L. Mitchell, *JPEG Still Image Data Compression Standard,* Van Nostrand Reinhold 1993.