Compression and Key Feature Extraction for Video Transmission

Esteban Tobias Bayro Kaiser¹, Eduardo Correa-Arameda¹,
and Eduardo Bayro-Corrochano²

¹ University of Tarapacá, Escuela universitaria de ingeniería eléctrica y electrónica
Arica-Chile

`locotoebk@hotmail.com, ecorrea@ut.edu.ch`

² CINVESTAV, Unidad Guadalajara, Departamento de Ingeniería Eléctrica y
Ciencias de la Computación, Jalisco, México
`edb@gdl.cinvestav.mx`

Abstract. This paper presents a gray scale image compression method using the Wavelet Transform and key feature detection for mobile phone video transmission. The major contribution of this work is to show the application of the wavelet transform in image compression and to add a new method to reduce redundant information in video transmission which is key feature detection. An algorithm is designed in Matlab to accomplish this task using a face to face video.

Keywords: Image processing, Compression, Wavelet transform, Key feature extraction, Face to face video.

1 Introduction

In the transmission of multimedia whether it is data, video or images, bandwidth and storage capacity are very important factors. Regardless of the constant improvement of storage devices and increasing the bandwidth, compression of information is still very necessary. There are several techniques for compressing information. These differ mainly in the amount of information that can be compressed and the amount of loss that is obtained when trying to restore them.

In this paper we will study the transmission of video between two people, this may be between two computers or between two cell phones; this type of transmission is known as face to face. The video that is used in this work consists of gray-scale images whose main feature is the face of the person. The images must be compressed or reduced in size to reduce the required bandwidth in the transmission [1].

For image compression there are several techniques available. In this paper the compression of the images is accomplished with the wavelet Transform.

In addition to analyse the image compression with the wavelet transform a method will be studied which transmits only the image key features [2]. These features are transmitted in sub-images. In a face to face transmission (for example: between cell phones), the key features are the most notable, and these are:

eyes and lips movements, therefore a technique is developed to locate and extract these features.

2 Compression of Images with the Wavelet Transform

Data compression refers to the process of reducing the volume of data needed to represent a certain amount of information, in this case gray level images. The compression is accomplished with the wavelet transform and Huffman coding. Huffman coding is an entropy encoding algorithm used for lossless data compression. The term coding refers to the use of a variable-length code table for encoding a source symbol where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol.

3 Key Features Extraction

The key features extraction of an image can be useful for reducing the amount of information in a video transmission. This information comes within a matrix, so if the size can be reduced, then the image has a representation with less information.

In a video each image compared with its previous suffers very little change, this feature can be harnesssed to just focus on the analysis and transmission of these. When the video comes from a conversation between two people focusing on the head with a single fund the major changes affecting the image are the eye and lip movements.

If the eyes and lips extraction is success, then the full image can be represented by a previous image and the extracted sub-image.

For a face to face video, the steps for key features extraction are:

- Eye position detection, Image alignment, Sub-images extraction and Difference estimation between images.

3.1 Eye Position Detection

For eye position detection there are several steps that are listed below.

Key features Position

In order to find key features in the image, we binarize the gray scale image by applying a Sobel filter and a thresholding to the edge image [3].

In Fig. 3.1 b) it can be observed that eyes and mouth are dominating features in the image, thus there can be found a method to obtain the position of these key features.

These positions can be found through horizontal and vertical projections, analyzing the maximum and minimum of these functions. Then it is possible to determine positions of certain features such as eyes, mouth, nose, upper and lower head and side edges. Assuming $I(x, y)$ be the intensity of a image, the

Fig. 3.1. a) Gray scale image, b) Edge detection with a Sobel filter and a $T = 0.5804$ thresholding

Fig. 3.2. a) Points where the horizontal and vertical projected are analyzed. b) Geometric relationship between eyes and mouth. c) Rotation angle.

equations are: $HI(x) = \sum_{x=1}^m I(x, y)$, $VI(y) = \sum_{y=1}^n I(x, y)$, where the HI corresponds to the horizontal and VI vertical projection (see Fig. 3.2 a)).

Geometric relationship between eyes and mouth

To obtain the position of the mouth, the geometric relationship between eyes and mouth can be used. Knowing that the distance between the eyes is equal to the distance from eye to mouth, it is possible to find the respective coordinates with an isosceles triangle on the face (see Fig. 3.2 b)).

3.2 Image Alignment and Rotation Angle

The alignment can be accomplished by rotation. To determine the rotation angle it is necessary to consider the following: If the goal is always to have the eyes lined up in a horizontal line then it is necessary to find the angle between the horizontal and eyes, as shown in Fig. 3.2 c)).

To find this angle requires the eyes coordinates, left eye represented by the point (x_{left}, y_{left}) and the right eye by (x_{right}, y_{right}) . The angle can be determined by $\alpha = \tan^{-1}(\frac{y_{right} - y_{left}}{x_{right} - x_{left}})$. It is worth mentioning that the image can be rotated both directions.

3.3 Sub-image Extraction and Difference Estimation between Images

Once the image has been aligned, the image can be divided into sub-matrices; as it used the coordinates of the eyes and mouth, the adequate estimated size

matrices will contain the entire feature. The centers of these sub-matrices are the coordinates for each feature.

To estimate the difference between each image or sub-image there must be a cost function, and the most appropriate is MAE (Mean Absolute Error). $MAE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |I(i, j) - I'(i, j)|$, where I is an image and I' is its previous one. The value delivered by MAE will be positive and real. To decide whether an image changes significantly, it is necessary to calculate the MAE and then find a value that can also be called thresholding, which will be called thresholding of difference T . This value will be the decision point if the sub-image is transmitted or not.

It's important to note that this value of T difference is calculated by considering the visual effect of change that each image has with the previous one. Then the following comparison must be done, if MAE is greater than the difference of T , the image is sent, otherwise there is no need to send it because the difference is very small and it is sufficient to represent the sub-image with its predecessor.

4 Code Results in Matlab

This section shows the analysis results of the algorithms written with Matlab software [4]. The algorithm can be divided into two parts: one considers its earlier phase of pre-transmission, and post-transmission. Both are shown in Figs. 4.1 a) and b).

4.1 Pre-transmission Algorithm

The pre-transmission phase will be put under performance test, to do this, the phase is divided into two parts that allow an individual analysis for each stage. Different types of face images are used, with different characteristics and situations, which helps to determine the functionality and efficiency of the algorithm.

4.2.1 Algorithm for the Extraction of Important Features and Decision Making

The Fig. 4.1 c) shows the algorithm to extract important features and decision-making. In this stage, the image is aligned and focused, and the algorithm is responsible for selecting those parts of the image that are relevant, such as detecting the movements of eyes and lips.

In this case, it is considered image patterns in gray level and taken in front of a single person. The background of these pictures should be simple and continuous, that is, without a background objects that change shape. In addition the person must be looking at the camera at all times.

For a better appreciation each block will be analyzed separately.

Image Analysis

The image size is 135 x 99 pixels.

Fig. 4.1. Results of the 3D reconstruction process. a) Pre-transmission algorithm. b) Post-transmission algorithm. c) Algorithm to extract important features and decision-making. Figures a) and b) show that every part of algorithm consists of two major sections, this is done in order to analyze the efficiency of the algorithm.

i) Edge detection. Figure 4.2 shows that the edges of the image are well defined, which indicates that there is no objection to proceed to the next step: the horizontal and vertical projections.

ii) Horizontal and vertical projections. In the horizontal projection of the Fig. 4.3 a) the first 2 maximum are obtained, these indicate the sides of the face. In the vertical projection of Fig. 4.3 b) the maximum value indicates the central line of the eye.

- Maximum horizontal projection: 26 and 77. Maximum vertical projection: 42.

With these points it is possible to obtain two sub-images where each one contains an eye. Then the same procedure is repeated.

iii) Sub-image Edge detection. Note: The left and right refer to as seen by the observer.

In Figs. 4.4 a) and b) the eye is identified easily without other information that could interfere with the projections, with this step any information that is not of interest was removed to find the exact iris coordinates.

iv) Sub-image horizontal and vertical projections.

- Maximum horizontal projection from the left eye: 17 (see Fig. 4.5). Maximum vertical projection from the left eye: 6.

Fig. 4.2. Edge detection with thresholding $T = 0.5608$

Fig. 4.3. a) Horizontal projection, b) Vertical projection

Fig. 4.4. Sub-image edge detection. a) left eye, with thresholding $T = 0.3980$. b) right eye, with thresholding $T = 0.4784$.

Fig. 4.5. a) Horizontal projection left eye and b) Vertical projection left eye

- Maximum horizontal projection from the right eye: 4 (see Fig. 4.6). Maximum vertical projection from the right eye: 12.

After finding these points, the offset that was created to generate the sub-image is added to obtain the iris location.

- Left eye position: (45, 43). Right eye position: (46, 64).

With these points and the geometric relationship between eyes and mouth, the mouth position can be calculated. Various tests determined that the distance from the center of the eye to the mouth is 1.2 D.

- Mouth position: (70, 52) (see Fig. 4.7).

The same procedure was carried out to various images and the algorithm was right with the coordinates of the features, as shown in Fig. 4.8.

Fig. 4.6. a) Horizontal projection right eye. b) Vertical projection right eye.

Fig. 4.7. Eyes and mouth position

Fig. 4.8. Eyes and mouth positions

5 Conclusions

One of the key factors that is analyzed in this work is to perform the transmission of video images, considering the least amount of information possible from each image, without significant loss of information, in an ideal channel. To achieve the image compression from a video using the wavelet transform an algorithm in Matlab was design resembling a JPEG200 compression [5]. The algorithm is able to compress and decompress images with the least information loss.

The images that were used for this algorithm allowed only the first level of compression, achieving a ratio of no more than 10:1 compression. The image compression that was achieved by extracting sub-images was about 90:1, this result shows that in the image compression key features extraction must be considered.

Acknowledgments. Esteban T. Bayro Kaiser, was supported by Msc Eduardo Correa Arameda and Prof. Eduardo Bayro Corrochano.

References

1. Ravyse, I., Sahli, H.: Facial analysis and synthesis. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2006. LNCS, vol. 4179, pp. 810–820. Springer, Heidelberg (2006)
2. Nixon, M., Aguado, A.: Feature Extraction and Image Processing, 1st edn. Elsevier Ltd., Amsterdam (2002)
3. Gonzalez, R., Woods, R.: Digital Image Processing using Matlab, 1st edn. Pearson Prentice hall, London (2004)
4. Mathworks: <http://www.mathworks.com/>
5. Usevitch, B.E.: A tutorial on modern lossy wavelet image compression: Foundations of jpeg 2000. IEEE Signal Processing Magazine 18, 22–35 (2001)