Expression Detector System based on Facial Images

José G. Hernández-Travieso, Carlos M. Travieso, Marcos del Pozo-Baños and Jesús B. Alonso
Signal and Communications Department, The Institute for Technological Development and Innovation on Communications, University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, sn, Ed. de Telecomunicación, Pabellón B, Despacho 111, E35017, Las Palmas de Gran Canaria, Spain

Keywords: Expression Detection, Soft-biometrics, Facial Segmentation, Pattern Recognition.

Abstract: This paper proposes a emotion detector, applied for facial images, based on the analysis of facial segmentation. The parameterizations have been developed on spatial and transform domains, and the classification has been done by Support Vector Machines. A public database has been used in experiments, The Radboud Faces Database (RAFD), with eight possible emotions: anger, disgust, fear, happiness, sadness, surprise, neutral and contempt. Our best approach has been reached with decision fusion, using transform domains, reaching an accurate up to 96.62%.

1 INTRODUCTION

In today's society, the use of Information and Communication Technologies (ICT) is increasing (Chin et al., 2008); (Eshete et al., 2010); (Siriak and Islam, 2010). Technological advances have made possible the proliferation of equipment and latest technologies, making progresses that could previously only imagine. One of many new applications is the emotion detection, being the goal of this work. It can be used for various purposes, as the detection of possible symptoms of neurological diseases in humans (Wang et al., 2008); (Wang et al., 2007); (Ekman and Friesen, 1978).

It is also gaining importance the Emotional Intelligence and another set of values and behaviours aimed at achieving better welfare of the individual in their work environment, emotional and affective.

This field of emotion detection is developing in multiple applications and researches, which gives an idea of the importance acquired and the multitude of applications thereof. In this regard, many authors are based on guidelines set by Ekman and Friesen, who developed the Facial Action Coding System (FACS) (Ekman and Friesen, 1978) that takes parameters of the muscles of the face according to a particular emotion, classifying them into Action Units (AU) specific to each emotion.

The use of FACS is not limited to the field of technological research, as it also has a great importance in helping psychology to study human behaviour. Only when an emotion is true, the correct AU is made, something that does not happens when you lie.

When transmitting a message, an important part of the communication is the facial expression, the gestures shown.

The state of the art in this field is quite broad, emphasizing at this point only a few jobs.

As mentioned before, the implementation of FACS has influenced works like (Pantic and Patras, 2004), who marked key points in the input images to the system to detect the presence of emotion. In this work, they found that the left half of the face expresses emotion better than the right half. In addition, it was found that the expression of authentic emotions were symmetrical, other than face feigned expressions. They used Hidden Markov Model (HMM) reaching recognition rates of 87%.

(Arima et al., 2004) using Fourier descriptors and discriminant analysis, studied the human response to low frequency oscillations using simulator ship movements and its passengers, to study the effect of the boat trip oscillations. They tried to establish a method of quantification of facial expression and clarify the relationship between facial expression and individual's mental status, managing to reach an average rate of recognition of 82.2%.

In (Wong and Cho, 2006), using Gabor features, developed a representation of facial emotion in Face Emotion Tree Structure (FEETS) to detect emotions

Hernández-Travieso J., Travieso C., del Pozo-Baños M. and Alonso J. (2013).
Expression Detector System based on Facial Images.
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*, pages 411-418
DOI: 10.5220/0004322504110418

in faces partially covered by sunglasses, veils, or any element that hides from view any area, achieving facial expression recognition results close to 90%.

(Fu et al., 2009) conducted a study which used Java Agent Development Framework (JADE), which linked the activity of the viewer using the remote control combined with facial recognition, with the emotions of the human being. Work that can be used to support the research of the Massachusetts Institute of Technology (MIT) on the home of the future, in which changes in the conditions of blood pressure, weight or abnormal sleep, are monitored as precursors of heart failure symptoms.

Also the study of (An and Chung, 2009) was carried out, using Principal Components Analysis (PCA) to study facial expression, when offering an interactive TV and, on demand, offering personalized services to viewers. In this study, they achieved a success rate of 92.1%.

In (Petrantonakis and Hadjileontiadis, 2010), using High Order Crossing Analysis (HOC), implemented an emotion detector system based on electroencephalogram (EEG), observing the graphs obtained by showing a facial expression of certain emotions. They achieved success rates of 100%.

(Dahmane and Meunier, 2001) developed an emotions detector system, using histograms of oriented gradients and Support Vector Machine (SVM) for the classification of images used, achieving a success rate of 70%.

Also (Gouizi, et al., 2001) developed an emotion detector system from biological signals such as electromyogram, respiration, skin temperature, skin conductance, blood pressure and rate pressure. SVM were used as a technique of classification. Recognition rates reached 85%.

This area has developed some works during the last years, and this work contributes to extend this line, showing our innovation. In particular, our work proposes the creation of an emotion detector system for facial images. For that, facial features will be extracted using spatial domains and transformed domains for subsequent classification using SVM. The distinctive part of this system is the segmentation of the image, performing a deep study that leads to obtain the significant value of each one when an emotion is detected. This has not been observed in previous studies.

2 PREPROCESSING

This section is composed by different steps in order to do easy our face segmentation. Those steps are, firstly, the face detection, after, a brightness adjusting and a high pass filter and finally, a process of binarization.

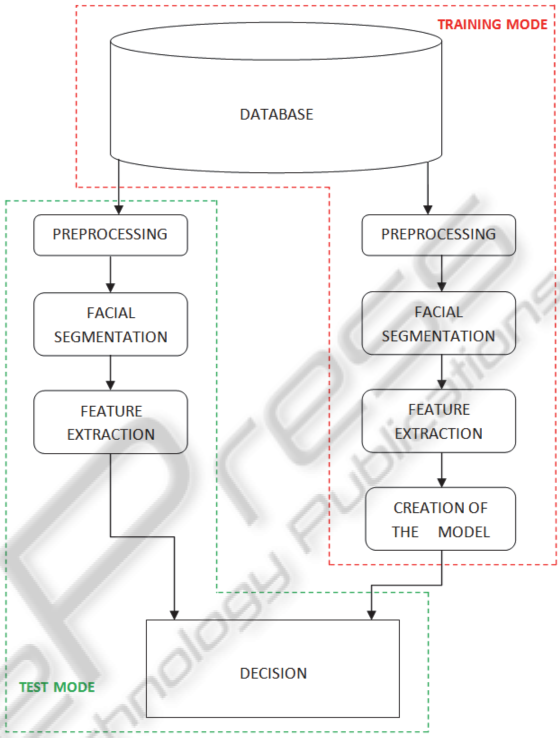

Figure 1: Block diagram of the system.

2.1 Extraction of the Facial Area of the Input Image

Due to the high resolution of the input images (681x1024 pixels) and that, in them, in addition to the facial area of interest, other body parts as the upper trunk and the top of the head are present, it proceeds to extract the facial area. An algorithm based on the face detector from (Viola and Jones, 2004), it has been used (see figure 2).

Figure 2: Extraction of facial area.

2.2 Adjusting the Brightness of the Image

The first step is to transform the incoming facial images to luminance and chrominance components, to highlight the eye and mouth areas of the face. Subsequently, the luminance component is used to modify the image brightness by multiplying the value component to be called ESCALA (see figure 3).

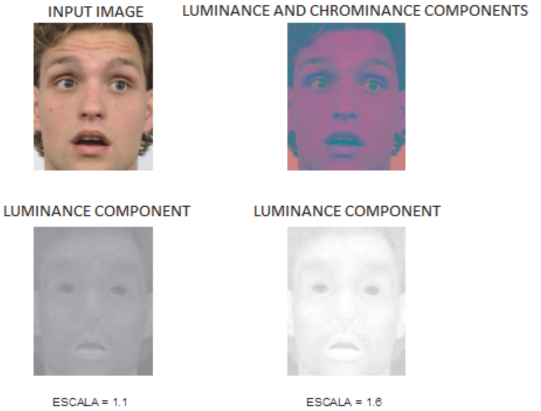

Figure 3: Adjusting the brightness.

2.3 Filtering of the Images

In order to obtain better information on areas of interest, to correctly detect the emotion present in the facial image, it requires a high pass filter for a better differentiation in the edges of the image (see figure 4). We have applied a heuristic filter, and finally, it is defined in the equation 1;

$$\mathrm{MFPA} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (1)$$

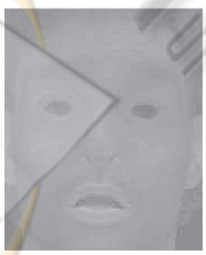

Figure 4: Filtered image.

2.4 Image Binarization

Binarization of the image consists in converting a gray scale image to a binary image, i.e., a black and white image. To binarize the image a histogram of the incoming picture luminance scale is made. It shows the maximum number of times that the values of the gray scale are present.

Using Otsu's Method (Otsu, 1979) is not feasible in this case, since the detection of valleys of the histogram is not optimal, shifting the threshold to lower values and losing information of important areas, such as, the mouth.

It then chooses a threshold manually, using the histogram, due to the need to find an optimal value for the parts involved in this study.

Figure 5: Binarized image.

3 FACIAL SEGMENTATION

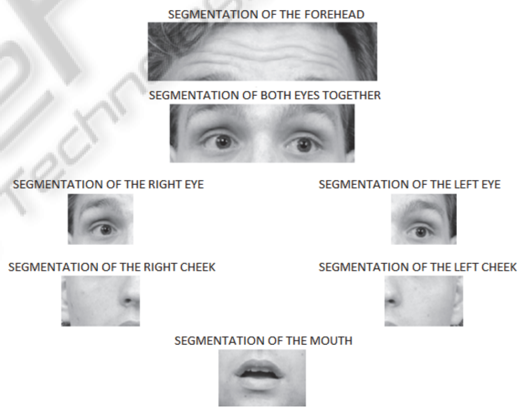

Figure 6: Segmented facial image.

Once the facial area has been pre-processed, it is segmented into seven parts to discern on the information given per each one in our process of emotion detection. The segmented parts are: forehead, both eyes together, right eye, left eye, right cheek, left cheek and mouth. And in particular, the definition of each segment is as follows:

- TP: indicates that all segments of the facial image (forehead, two eyes together, right eye, left eye, right cheek, left cheek and mouth) are used.
- DOLOBO: indicates that both eyes together, right eye, left eye and mouth are used.
- DOBO: indicates that both eyes together and

mouth are used.

- LOBO: indicates that right eye, left eye and mouth are used.
- FR: indicates that forehead is used.
- DO: indicates that both eyes are used together.
- LO: indicates that right eye and left eye are used.

4 FEATURE EXTRACTION

4.1 Facial Feature Extraction in the Spatial Domain

The facial feature extraction in the spatial domain consists of taking Euclidean distances between various points of the face, with the binarized images, to try to detect and emotion present on it. These distances are normalized with respect to the distance between the inner ends of the eyes, due to the variety of the faces of the database for men, women and children, to try to standardize the measures taken.

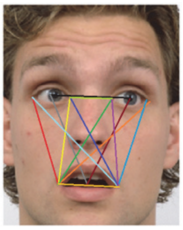

Figure 7: Euclidean distances.

4.2 Facial Feature Extraction in Transformed Domains

For this work, the used transformed domains are 2 Dimensional Discrete Cosine Transform (2D-DCT) (Gonzalez and Woods, 2002) and 2 Dimensional Discrete Wavelet Transform (2D-DWT). They were chosen due on their good behaviour in facial identification and other biometric applications (Vargas et al., 2010); (Fuertes et al., 2012).

On the one hand, the input image that will serve to 2D-DCT is high pass filtered. This process is performed to obtain a better definition of the edges of the image, achieving a better highlight area of facial expression characteristics such as, eyes and mouth, for later extraction. That information on the details, obtained filtering, is achieved through spectral windows, given that working in space-frequency resolution, that information must be transformed into the spatial domain. The 2D-DCT performs a low pass filter that provides general information from the details of the incoming image. On the other hand, the 2D-DWT (Gonzalez and Woods, 2002), carries a high pass filter which provides detailed information of the details from the incoming image, the image used in this case is the original image in colour (RGB). Being the image in the visible domain, spatial information is provided.

5 CLASSIFICATION SYSTEM

5.1 Support Vector Machine (Svm)

The SVM is a well-known classifier and used on different examples with large size of data (Yu et al., 2003). The SVM only can distinguish between two different classes (Vapnik, 1998); (Burges, 1998). The technique is directly related to classification and regression models (Vapnik, 1998). Given a set of training examples (samples, called vectors), can be labelled classes and train a SVM to build a model that predicts the kind of a new sample. The idea underlying the SVM is the hyperplane or decision level, which can be defined as the plane of separation between a set of samples from different classes. Hyperplanes can be infinite, but only one of them is the optimal one, this is what makes that the separation between the samples is maximized (Vapnik, 1998); (Jakkola, 2002) causing the margin is maximized. We have used a supervised classification system, with two different kernels, Linear and Radial Basis Function (RBF) kernels (Vapnik, 1998), under a one-versus-all multi-classes strategy. In particular, we have used a SVM-Light (Joachims, 1999).

5.2 Fusion Results

The last stage is the fusion of classification results. This fusion is at the decision level from the output of the SVM decision. Its mission is to correct certain errors, since they are uncorrelated, which may have occurred in the recognition phase. The objective proposed, is to give more robustness to the final results of our approach.

6 EXPERIMENTAL METHODOLOGY

6.1 Database

We used a public database, The Radboud Faces

Database (RAFD) (Langner et al., 2010). This database was chosen over other available by several factors, among which the brightness and image resolution.

The RAFD database is a set of 8040 images of 67 models (20 adult Caucasian male, 19 female Caucasian adults, 4 Caucasian children, 6 girls Caucasian, 18 Moroccan adult male) with 23-24 pictures per model for each position on the camera, which express 8 emotional expressions: anger, disgust, fear, happiness, sadness, surprise, neutral and contempt. Emotions expressed according to FACS. The positions of the models to the chamber range from -90° to 90° from the front of the camera (which is assumed 0°). The database is an initiative of the Institute of Behavioural Science of the Radboud University Nijmegen, located in Nijmegen (The Netherlands).

The file format is .jpg in colour, with dimensions of 681x1024 pixels. The clothing of all models is identical, a black shirt and the background is clear and unchanged. To this system, 1600 images corresponding to the front position of the model about the camera were used. This database is public, and is granted free of charge for use in research.

Figure 8: Samples of the database RAFD.

6.2 Experiments

In the experiments used SVM with RBF kernel and linear following 50% hold-out validation method, repeating the experiments three times, varying percentages of training and test samples.

Being originally a SVM bi-classes and working this system with more than two classes of emotions (8 in total), it requires a multiclass system. Among the several existing techniques, the one-versus-all technique was used. Two experiments have been developed;

6.2.1 Experiment 1: Feature Extraction in the Spatial Domain

In the case of facial feature extraction in the spatial domain, the distance measurements are concatenated into a column vector, and subsequently are introduced into a data structure, which will be the input data to the classification stage.

6.2.2 Experiment 2: Feature Extraction in Transformed Domains

The 2D-DCT applies to segmented images of the high pass filtered facial area. This transform has the property that the images do not undergo any variation in size to perform it.

Each time a segment has been transformed, becomes the data matrix which is formed in a column vector, then concatenating each column vector of each segment face to form a new column vector that is introduced into a data structure, which will be the input data to the classification stage.

In applying the 2D-DWT, followed a similar pattern to those followed in the 2D-DCT. In this case, the input image is high pass filtered, but is the original image, because the 2D-DWT works with images in colour (RGB).

Among the different types of existing wavelet, we chose to use the Haar family for its simplicity and family Bior4.4 due to its good result in previous works (Mallat, 2009).

From the application of 2D-DWT, we have worked with the high frequency, in order to get the details of each image. This output image becomes a column vector, as occurs with the 2D-DCT, by concatenating all column vectors in columns corresponding to the selected facial segments to form a new column vector, that is introduced into a data structure and it will be the input data to the classification stage.

6.2.3 Experiment 3: Fusion

Once we have obtained the simulation results for facial feature extraction in transformed domains (2D-DCT, Bior4.4 2D-DWT and Haar 2D-DWT), a fusion of the best results from each are performed. This is achieved uncorrelated correct errors and improves the emotion recognition.

6.3 Results and Discussion

The results are shown in mean and variance for each of the experiments performed.

6.3.1 Experiment 1: Feature Extraction in the Spatial Domain

The best result obtained using the spatial domain was of 32.58% ± 1.00 with a 50% of training samples and using linear SVM.

In view of these results, it is proved that this method is not decisive for detecting an emotion present in the human being using the facial image, because the information is not sufficient to achieve a percentage of recognition enabling determine with certainty the emotion present in the facial image.

6.3.2 Experiment 2: Feature Extraction in Transformed Domains

By employing transformed domains was obtained the following results, for 50% of training samples;

- For 2D-DCT, it was 96.16% ± 0.69 with RBF SVM, using TP and 86.41% ± 1.34 with Linear SVM using DOLOBO.
- For the case of Haar 2D-DWT, the best result obtained was 86.41% ± 1.34 for RBF SVM using TP, and 92.90% ± 0.33 for Linear SVM using TP.
- For Bior4.4 2D-DWT, 96.33% ± 1.34, using RBF SVM for TP, and 92.95% ± 1.97 for linear SVM using TP.

If 60% of test samples were used, the results were;

- For 2D-DCT was 90.52% ± 0.09 for RBF SVM using TP, and 87.77% ± 0.22 for linear SVM using TP.
- With Haar 2D-DWT, 94.70% ± 0.62 for RBF SVM using DOBO, and 91.18%±0.06 for Linear SVM using TP.
- For 2D-DWT bior4.4, 96.59% ± 0.32 for RBF SVM using LOBO, and 92.46% ± 2.73 for Linear SVM using TP.

In view of these results, it will conclude that the extraction of facial features in transformed domains is more effective to detect the emotion present in the facial image of the human being. The most effective one is Bior4.4 2D-DWT.

6.3.3 Experiment 3: Fusion

The best result for each percentage linear SVM is chosen for fusion, the result obtained for 50% of samples test was 96.62% success rate with a time of 28.86 milliseconds.

For 60% of test samples, the result obtained was 95.72% success rate with a time of 23.80 milliseconds.

With these values, it is clear the improvement experienced in applying fusion for detecting the emotion present in the human being.

Compared to previous systems in which there has been no segmentation for detecting emotion, this study achieved success rates over them. Thus it proves the advantage of segmentation to detect correctly the emotion present. Nowadays, The RAFD Face Database has not been used to detect emotions.

Table 1: Spatial domain results.

Spatial Domain Results	SVM	
	Linear	*RBF*
50% training	32.58% ± 1.00	25.41% ± 0.41
40% training	32.11% ± 0.02	22.77% ± 2.13

Table 2: 2D-DCT results.

Transformed Domain Results	2D-DCT	
	Linear SVM	***RBF SVM***
50% training (type of segment)	86.41% ± 1.34 (DOLOBO)	96.16% ± 0.69 (TP)
40% training (type of segment)	87.77% ± 0.22 (TP)	90.52% ± 0.09 (TP)

Table 3: Haar 2D-DWT results.

Transformed Domain Results	Haar 2D-DWT	
	Linear SVM	***RBF SVM***
50% training (type of segment)	92.90% ± 0.33 (TP)	95.37% ± 1.82 (TP)
40% training (type of segment)	91.18%±0.06 (TP)	94.70% ± 0.62 (DOBO)

Table 4: Bior4.4 2D-DWT results.

Transformed Domain Results	Bior4.42D-DWT	
	Linear SVM	***RBF SVM***
50% training (type of segment)	92.95% ± 1.97 (TP)	96.33% ± 1.34 (TP)
40% training (type of segment)	92.46% ± 2.73 (TP)	96.59% ± 0.32 (LOBO)

7 CONCLUSIONS

Once realized the study, it has shown that the segmentation of the face, its parametrization with transform domains and the use of SVM classifier gives a much higher percentage of recognition in simulations with transformed domains in the segments of the eye (in whole or separately) and the mouth are present together, reaching accurate of 96.59% using RBF SVM and 2D-DWT bior4.4.

In contrast, the less influential zones on the detection of emotion are the cheeks and forehead, due to the limited amount of information being given. Especially, the forehead, the results were not higher to 33.33%, using in this case, the Haar wavelet family.

The importance of the information provided by eyes and mouth is also checked empirically, because when a person shows emotions, like surprise, the parts of the face that more quickly and clearly serve as indicative are the eyes and mouth. By showing the eyes and mouth wide open, the emotion can be detected without any doubts. Which does not occur with the cheeks and forehead if considered separately, because the movements of the muscles associated with these areas is inconclusive in this study.

ACKNOWLEDGEMENTS

This work is partially supported by funds from "*Cátedra Telefónica 2009/10–ULPGC*" and by the Spanish Government, under Grant MCINN TEC2012-38630-C04-02.

REFERENCES

An, K. H., Chung, M. J., 2009. Cognitive face analysis system for future interactive TV. In *IEEE Transactions on Consumer Electronics*. Vol. 55, no. 4, pp. 2271-2279.

Arima, M., Ikeda, K., Hosoda, R., 2004. Analyses of Facial Expressions for the Evaluation of Seasickness. In *Oceans '04. MTTS/ IEEE Techno-Ocean '04*. Vol.2, pp. 1129-1132.

Burges, C. J. C., 1998. A tutorial on Support Vector Machines for Pattern Recognition. In *Data Mining and Knowledge Discovery*, Vol. 2, pp.121-167.

Chin, K. L., Chang, E., Atkinson, D., 2008. A Digital Ecosystem for ICT Educators, ICT Industry and ICT Students. In *Second IEEE International Conference on Digital Ecosystems and Technologies*. pp. 660-673.

Dahmane, M., Meunier, J., 2011. Emotion Recognition using Dynamic Grid-based HoG Features. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, pp. 884-888.

Ekman, P., Friesen, W., 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movements*. Consulting Psychologist Press, Palo Alto, CA.

Eshete, B., Mattioli, A., Villafiorita, A., Weldemariam, K., 2010. ICT for Good: Opportunities, Challenges and the Way Forward. In *Fourth International Conference on Digital Society*. pp. 14-19.

Fu, M. H., Kuo, Y. H., Lee, K. R., 2009. Fusing Remote Control Usage and Facial Expression for Emotion Recognition. In *Fourth International Conference on Innovative Computing, Information and Control*. pp. 132-135.

Fuertes, J. J., Travieso, C. M., Naranjo, V., 2012. *2-D Discrete Wavelet Transform for Hand Palm Texture Biometric Identification and Verification*. Wavelet Transforms and Their Recent Applications in Biology and Geoscience, Ed. InTech.

González, R. C., Woods, R. E., 2002. *Digital Image Processing*. Prentice Hall, Upper Saddle River, New Jersey.

Gouizi, K., Reguig, F. B., Maaoui, C., 2011. Analysis Physiological Signals for Emotion Recognition. In *7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*. pp. 147-150.

Jakkula, V., *Tutorial on Support Vector Machine (SVM)*. School of EECS, Washington State University, Pullman 99164.

Joachims, T., 1999. *Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning*. B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A., 2010. Presentation and validation of the Radboud Faces Database. In *Cognition & Emotion*, Vol. 24, nº 8, pp. 1377-1388.

Mallat, S., 2009. *A Wavelet Tour of Signal Processing*. Third Edition: The Sparse Way.

Otsu, N., 1979. A Threshold Selection Method from Gray_Level Histograms. In *IEEE Transactions on Systems, Man and Cybernetics*. Vol. 9, no. 1, pp. 62-66.

Pantic, M., Patras, I., 2006. Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments From Face Profile Image Sequences. In *IEEE Transactions on System, Man and Cybernetics-Part B: Cybernetics*. Vol. 36, no.2, pp. 443- 449.

Petrantonakis, P. C., Hadjileontiadis, L. J., 2010. Emotion Recognition from EEG Using High Order Crossing. In *IEEE Transactions on Information Technology in Biomedicine*. Vol. 14, no. 2, pp. 186-197.

Siriak, S., Islam, N., 2010. Relationship between Information and Communication Technology (ICT) Adoption and Hotel Productivity: An Empirical Study of the Hotels in Phuket, Thailand. In *Proceedings of PICMET'10: Technology Management for Global Economic Growth*, pp. 1-9.

Vapnik, N. V., 1998. *Statistical Learning Theory*. Wiley Interscience Publication, John Wiley & Sons Inc.,

Vargas, J. F., Travieso, C. M., Alonso, J. B., Ferrer. M. A., 2010. Off-line signature Verification Based on Gray Level Information Using Wavelet Transform and Texture Features. In *12th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. pp. 587-592.

Viola, P., Jones, M. J., 2004. Robust Real-Time Face Detection. In *International Journal of Computer Vision*. Vol. 57, nº 2, pp. 137-154.

Wang, P., Kohler, C., Barrett, F., Gur, R., Verma, R., 2007. Quantifying Facial Expression Abnormality in Schizophrenia by Combining 2D and 3D Features. In *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1-8.

Wang, P., Kohler, C., Martin E., Stolar, N., Verma, R., 2008. Learning-based Analysis of Emotional Impairments in Schizophrenia. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. pp. 1-8.

Wong, J. J., Cho, S. Y., 2006. Recognizing Human Emotion from Partial Facial Features. In *International International Joint Conference on Neural Networks*. Vol. 1, pp. 166-173.

Yu, H., Yang, J., Han, J., 2003. Classifying Large Data Sets Using SVMs with Hierarchical Clusters In *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 306-315.