Preliminary steps towards the bio-inspired detection of deceitful clues in facial expressions

Manuel HERNANDEZ HERNANDEZ,
Pedro Luis SANCHEZ ORELLANA,
Claudio CASTELLANOS SANCHEZ
Laboratory of Information Technology
Cinvestav-Tamaulipas, Ciudad Victoria, Tamaulipas, Mexico
Email: {mhernandez,psanchez,castellanos}@tamps.cinvestav.mx

Luis Carlos ORTEGA TAMEZ
Hospital Infantil de Tamaulipas
Ciudad Victoria, Tamaulipas, Mexico
Email: luiscarlos@cenepi.com

***Abstract*—We present the first steps of a bio-inspired model for detection of deceitful clues in facial expressions to clustering in suspects and non suspect persons. After build an anthropomorphical grid on the face, a convolution inspired in the neurons of V1-MT allows to obtain the active neuron ratios and the total energy of active neurons for each region in our anthropomorphical grid. The classification in suspect/non-suspect emerge from the difference between ratios of symmetrical regions in the face.**

I. Introduction

Over the past years there has been an increasing surge of interest in automated facial expression analysis. Gesture recognition can be termed as an approach in this direction. But the facial asymmetry (with left face part stronger than right one part) is apparent only with deliberate and not spontaneous (non natural) expressions [1].

In the literature the analysis of the facial expressions can be achieved by two main approaches [2]: Analytic and Holistic. In the first one flexible mathematical models are used to incorporate face deformation and illumination changes. In the holistic approach generally a grey-level template of the face is used to globally compare with a previously learned. On one hand, the analytical approach usually deals with the geometry of the face, proposing method for both 2D and 3D facial region detection [3]. On the other by considering the whole face as features [4]. Besides, the main emphasis is put on the classifier which has to deal with multiple faces in conditions different to the existing during training.

In spite of the proposed techniques and the results achieved there are several situations where the analysis of the expressions can not be successfully done by neither the analytic nor the holistic approaches due to the variations in the environment conditions (illumination changes). To solve this situation one can look at systems that efficiently locate, extract and analyse the information of the face, an example of this systems is the brain of the humans. According to several researchers [5]–[7] the processing of the emotions is done in the amygdala, an area located within medial temporal lobes in the brain. Image analyse obtained by fMRI [8]–[10], in this area the processing is modulated by the affective significance of faces, particularly with fearful expressions, but also with other social signals such gaze direction. Visual search paradigms have provided evidence for the enhanced capture of attention by threatening faces. Experiments carried out by Weymar et al. [11] demonstrated that the recognition process can be achieved with the single facial features (eyebrows and eyes vs. eyebrows) of threatening and friendly faces. This experiments confirm that emotion processing involves other areas like the visual cortex (e.g., V1) [8].

Neurophysiological studies [12], [13] have shown that some neurons in the superior temporal sulcus and the inferior temporal gyrus of macaque monkeys respond to faces. Neurons respond mainly in two separated ways: on one hand, to expression primarily in the cortex in the superior temporal sulcus (which involves changeable aspects of the face). On the other hand, neurons respond to identity mainly in the inferior temporal gyrus (analysis of the invariant aspects of the face), both modulated by attentional mechanisms [14]. Later, Hung et al in 2010 [15] demonstrated that fast, automatic, and parallel processing of unattended emotional faces, provides important insights into the specific and dissociated neural pathways in emotion and face perception [15].

Although, the areas involved in the analyse of facial expressions in human brain are clear, there is no consensus about how this process is done. Clues from the psychological point of view, like the proposed by Ekman and Friesen [16], that remarks the importance of the symmetry on this process. According to this work, the symmetrical expressions show a true (natural) emotion. Our research takes this affirmation for lead us to introduce our proposed architecture based on detection of asymmetrical face expressions, next, in their clustering in suspect/non-suspect deceitful clues persons. In the following we will present an architecture to analyse the asymmetrical responses of the face from a sequence of images. We mention the experiments and discuss the results to finally comment the future work.

II. Proposed bio-inspired architecture

A general diagram of the methodology is shown in the figure 1. Next subsections we give a brief explain the steps of our approach where we assuming that the face has been located correctly.

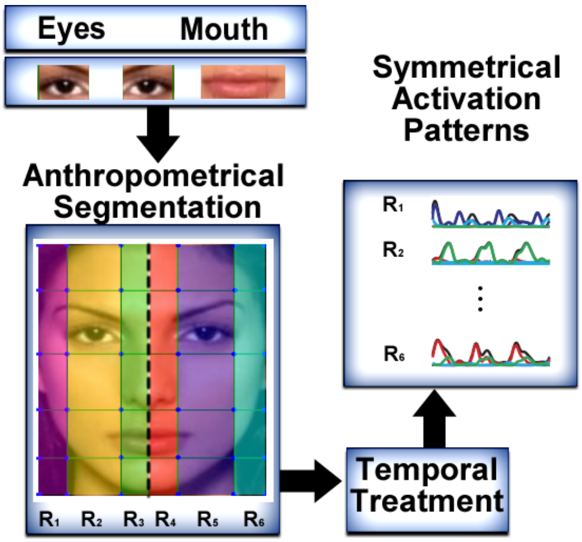


Fig. 1. General architecture of bio-inspired proposed model.

A. Anthropomorphical grid

To get the grid we working on the rectangle that contain the face detected. So let $I = \{I_1, I_2, I_3, \ldots I_t\}$ be a sequence of images of the face where the eyes and mouth corners are correctly detected. The coordinates of the corners of eyes and mouth located are used to segment the face into small regions by using anthropometrical measurements of face.

B. Bio-inspired processing

So considering the processing done by the visual cortex the first neurons that receive the stimuli from the eyes are the simple cells in primary visual cortex. The behaviour of these neurons show a preference to specific orientations, which computationally can be modelled by the oriented Gabor-like filters:

$$S_\theta(x,y,t) \quad = \quad \sum\sum_{t=0} I(x,y,t) * G_\theta(\hat{x},\hat{y}) \tag{1}$$

where $S_\theta(t)$ are simple cells which result of the convolution between Gabor functions ($G_\theta(\hat{x},\hat{y})$) and the image ($I(x,y,t)$), $\hat{x}$, $\hat{y}$ are the rotational components. The responses are then integrated by using a non-linear model that allows to merge the responses from the two phases. The complex cells responses are estimated by

$$C_\theta(x,y,t) = \sqrt{S_{\theta,\frac{\pi}{2}}(x,y)^2 + S_{\theta,(-\frac{\pi}{2})}(x,y)^2} \tag{2}$$

where $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ are the symmetric and anti-symmetric phases.

In the brain, this neurons are connected to MT neurons allowing a temporal processing that is defined as the temporal difference between $C_\theta(x,y,t)$ and $C_\theta(x,y,t-1)$ obtaining the active neurons by

$$D_\theta(x,y,t) = C_\theta(x,y,t) - C_\theta(x,y,t-1) \tag{3}$$

TABLE I
EMPIRICAL DATA FOR DETERMINING κ INDEX FOR TWO SECONDS. PERSON CONDITIONS (PC), HEAD MOTION (HM), AND LUMINANCE-CAPTURE (LC) CONDITIONS.

PC	HM	LC	κ index
3	1	1	
3	2	2	
3	3	3	PC+HM+LC
3	4	4	
3	5	5	

where $D_\theta(x,y,t)$ is the result of active neurons and we use this information to compute the number and the energy of active neurons.

C. Compute of the intersection energy and number of active neurons

Then for each region we compute the rate between the energy of active neurons and the maximum energy of neurons in the region. So also we compute the rate between the number of active neurons and the maximum number of active neurons. These are estimated by

$$\vec{R} = \left[\frac{C_{A_1}}{C_{E_1}}, \frac{C_{A_2}}{C_{E_2}}, \frac{C_{A_3}}{C_{E_3}}, \ldots, \frac{C_{A_r}}{C_{E_r}}\right]^T \tag{4}$$

where $\vec{R}$ is the vector of the rate for the regions r, C_{A_r} is the number or the energy of active neurons in the region r, C_{E_r} is the total number or the maximum energy of expected active neurons for each region r. Next, we compare each symmetrical region both the energy and number of active neurons with a threshold κ obtained empirically for two seconds and fix in the table I.

D. Clustering in suspect/non-suspect clusters.

For each sequence we obtain a kappa index adapted to numbers of images in the sequence.

$$C_\kappa = \frac{C_A}{C_E}\frac{2F}{T\kappa} \tag{5}$$

where $\frac{C_A}{C_E}$ is the ratio of energy/number of active neurons in a region, T is the total numbers of images in the sequence, F is the capture frequency for this sequence.

With C_κ values we clustering in: suspect, doubt and non-suspect clusters.

Finally, we intersect the both asymmetries detected: energy and number of active neurons. The results of this intersection show the regions where the facial expression is asymmetric.

III. RESULTS

A. Description of databases

We used three databases for test our proposed approach: FG-Net, CK+, and LTI-HIT.

Facial Expressions and Emotions from the Technical University Munich created in 2006 the FG-Net database. This one is an image database containing face images showing a number of subjects performing six different basic emotions defined by Ekman & Friesen: anger, fear, surprise, disgust, sadness and happiness, also they add the neutral expression for each person. This database consists of 19 persons with 21 sequences for each person, then it contain 399 different sequences. The sequences of images were captured to 25 frames per second with a resolution of 640×480 pixels [17]. This database allows to observe people react as natural as possible. As a consequence, it was tried to wake real emotions by playing video clips or still images after a short introduction phase instead of telling the person to play a role. The background of the original images rather than the part of the image including the face was used. This is in contrast to the asymmetry of luminance of the face, which includes not only asymmetry introduced by lighting, but also asymmetry introduced by the face itself.

The Institute of Electrical & Electronics Engineers at Grenoble, France, created in 2000 and 2010 the CK and CK+ (the same CK database but 107 new sequences of which 33 are in color) consists of 123 persons with 7 different expressions per person: anger, contempt, disgust, fear, happy, sadness and surprise. The image sequence vary in duration (i.e. 10 to 60 frames) and incorporate the onset (which is also the neutral frame) to peak formation of the facial expressions. This database was captured to 30 frames per second with a resolution 640×480. The final frame of each image sequence was coded using FACS (Facial Action Coding System) which describes subject's expression in terms of action units (AUs) [18]. Participants were instructed by an experimenter to perform a series of 23 facial displays included single action units and emotion-specified expressions of joy, surprise, anger, disgust, fear, and sadness. Each display began and ended in a neutral face with any exceptions noted.

Finally, the Children Hospital of Tamaulipas created in 2010 the LTI-HIT. This database is issue of interviews that consists of 52 persons (11 men and 41 women) into 2 different environments (5 and 47 persons) with 30 different questions. Each sequence of images were captured to 30 frames per second with to resolution 720×480 pixels and duration of 1.6 to 2.5 minutes per interview. The interviews were make in natural conditions.

B. Data preparation

For tests with our approach we take only the sequence of images to colour for the CK+ database while for FGNet and LTI-HIT we take all sequences images.

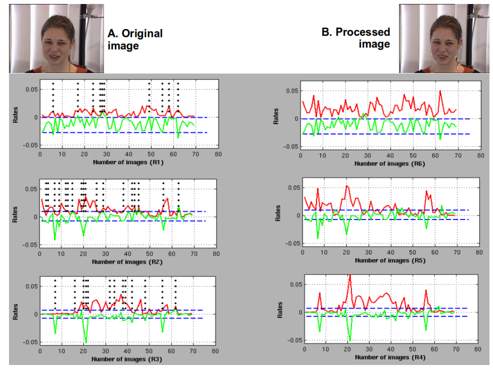
C. Results

We take six columns of proposed face grid. The symmetrical regions are: R1 with R6 (ears), R2 with R5 (eyes) and R3 with R4 (nose) (figure 1).

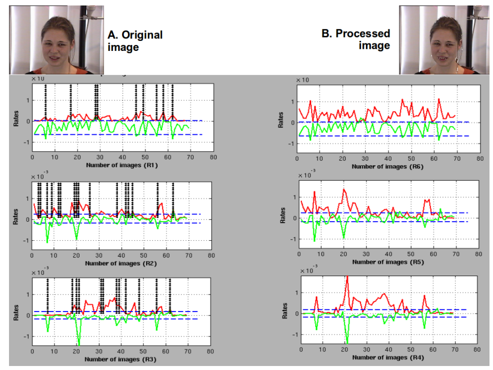
The three databases are very different in its acquisition, illumination and manipulation to capture conditions. For compare these ones, we propose the κ index according to table I.

In natural conditions, left-sided change of face is only about 2% greater than right-sided change overall [19]. We stablish a level of 1 to 5 for the person conditions and, in average, we choice a level 3 for model our all databases.

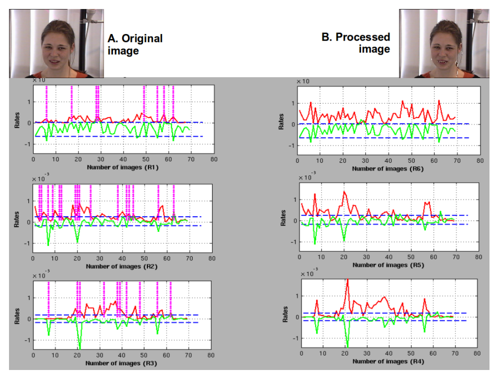
The head motion is very different too between these three databases. In CK+ all persons are very controlled and the original sequences were cut for only show the evident expression. Against, the FG-Net sequences show before and after the evident expression. Finally, the LTI-HIT sequences, was not



(a) Quantity of active neurons



(b) Energy of active neurons



(c) Quantity/energy intersection of active neurons

Fig. 2. We apply our approach for asymmetries detection in the face.

TABLE II
ASSIGNED κ INDEX TO OUR THREE DATABASES.

Database	PC	HM	LC	κ index
CK+	3	1	2	6
FG-Net	3	3	3	9
LTI-HIT	3	4	4	11

TABLE III

SUSPECT/NON-SUSPECT RATIOS IN THE DATABASE SEQUENCES ACCORDING κ INDEX FOR THREE MANIPULATIONS: QUANTITY OF ACTIVE NEURONS (QAN), ENERGY OF ACTIVE NEURONES (EAN), COMBINATION RATIO BETWEEN QUANTITY AND ENERGY OF ACTIVE NEURONS (IQEAN).

		Non-susp			Suspect	
DB	QAN	EAN	IQEAN	QAN	EAN	IQEAN
CK+	0.303	0.242	0.333	0.636	0.636	0.546
FG-Net	0.243	0.286	0.579	0.476	0.376	0.109
LTI-HIT	0.177	0.242	0.552	0.576	0.499	0.153

prepare for the major expression. It shows the natural reactions of interviewer persons. Then, we assign to HM variable 1, 3 and 4 to each database, respectively.

The luminance and capture conditions is also too very different. CK+ was created in a studio, FG-Net in a controlled environment and LTI-HIT in a uncontrolled environment. Then, we assign to LC variable 2, 3 and 4 to each database, respectively.

The table II resume the values assigned of κ index to three databases.

We test our approach for quantity of active neurons, total energy of active neurons and ratio combination of these first two ones. The figure 2 show an example of tests for the sequences of images of the three databases.

The table III resume the obtained ratios for suspect/non-suspect persons and for each database. The QAN and EAN ratios for non-suspect persons to deceitful clues is always minor that suspect persons for all tested databases. But, the combination ratio of QAN and EAN (IQEAN), the databases FG-Net and LTI-HIT show major non-suspect persons that suspect persons. These observations is consistent with the found characteristics in the literature of each database.

The figure 3 show the clustering of three tested databases. In this figure, the CK+ database show a ratio of suspect persons major that the non-suspect persons. Against, the other two ones, the major is the ratio of non-suspect persons.

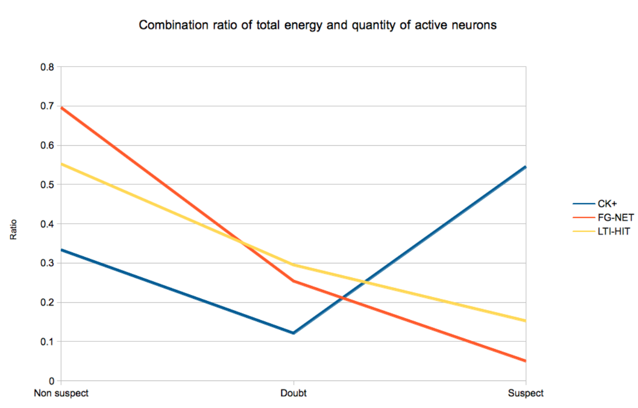


Fig. 3. Distribution of suspect/non-suspect persons for each tested databases.

IV. CONCLUSIONS AND FUTURE WORK

We shown a bio-inspired approach to detection of deceitful clues in facial expressions. After face detected, an anthropomorphical grid is proposed. For each region, we obtain the ratio of quantity/energy of active neurons issues of the modelled simple and complex neurons by our Gabor-like filters.

The different conditions of each database not allow a simple characterization. Then, we proposed an experimental index for modelled these conditions, κ. In the future, we will determine automatically this index.

But, our preliminary steps of our approach show the feasibility to detect suspect persons: nervous or alteration situations and lie's actions. This is our beginning to deceitful clues in facial expressions.

ACKNOWLEDGMENT

This research was partially funded by project number 51623 from "Fondo Mixto Conacyt-Gobierno del Estado de Tamaulipas" and the project number 78885 from "Ciencia Basica 2007" of Conacyt.

REFERENCES

[1] P. Ekman, "Asymmetry in facial expression," *Science*, vol. 209, no. 4458, pp. 833–834, 1980.

[2] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man and Cybernetics - Part C*, vol. 37, no. 3, pp. 311–324, 2007.

[3] A. Franco, D. Maio, and D. Maltoni, "2d face recognition based on supervised subspace learning from 3d models," *Pattern Recognition*, vol. 41, no. 12, pp. 3822 – 3833, 2008.

[4] D. Delgado-Gomez, J. Fagertun, B. Ersbøll, F. M. Sukno, and A. F. Frangi, "Similarity-based fisherfaces," *Pattern Recogn. Lett.*, vol. 30, pp. 1110–1116, September 2009.

[5] M. L. Phillips, A. W. Young, C. Senior, M. Brammer, C. Andrews, A. J. Calder, E. T. Bullmore, D. I. Perrett, D. Rowland, S. C. R. Williams, J. A. Gray, and A. S. David, "Aspecific neural substrate for perceiving facial expressions of disgust," *Nature*, vol. 389, no. 6650, pp. 174 – 194, 1997.

[6] R. Adolphs, D. Tranel, H. Damasio, and A. R. Damasio, "Fear and the human amygdala," *The Journal of Neuroscience*, vol. 15, no. 9, pp. 5879–5891, 1995.

[7] P. J. Whalen, S. L. Rauch, N. L. Etcoff, S. C. McInerney, M. B. Lee, and M. A. Jenike, "Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge." *The Journal of Neuroscience*, vol. 18, no. 1, pp. 411–418, January-1998.

[8] P. Vuilleumier and G. Pourtois, "Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging," *Neuropsychologia*, vol. 45, no. 1, pp. 174 – 194, 2007.

[9] R. Adolphs, D. Tranel, and A. R. Damasio, "The human amygdala in social judgment." *Nature*, vol. 393, no. 6684, pp. 470–474, Sep-1998.

[10] F. Gosselin, M. L. Spezio, D. Tranel, and R. Adolphs, "Asymmetrical use of eye information from faces following unilateral amygdala damage," *Social Cognitive and Affective Neuroscience*, vol. 6, no. 3, pp. 330–337, 2011.

[11] M. Weymar, A. Law, A. Ohman, and A. O. Hamm, "The face is more than its parts brain dynamics of enhanced spatial attention to schematic threat," *NeuroImage*, vol. 58, no. 3, pp. 946 – 954, 2011.

[12] M. E. Hasselmo, E. T. Rolls, and G. C. Baylis, "The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey," *Behavioural Brain Research*, vol. 32, no. 3, pp. 203 – 218, 1989.

[13] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, "The distributed human neural system for face perception." *Trends in cognitive sciences*, vol. 4, no. 6, pp. 223–233, 2000.

[14] M. A. Williams, F. McGlone, D. F. Abbott, and J. B. Mattingley, "Differential amygdala responses to happy and fearful facial expressions depend on selective attention," *NeuroImage*, vol. 24, no. 2, pp. 417 – 425, 2005.

[15] Y. Hung, M. L. Smith, D. J. Bayle, T. Mills, D. Cheyne, and M. J. Taylor, "Unattended emotional faces elicit early lateralized amygdala-frontal and fusiform activations," *NeuroImage*, vol. 50, no. 2, pp. 727 – 733, 2010.

[16] P. Ekman and W. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Palo Alto: Consulting Psychologists Press, 1978.

[17] F. Wallhoff, "Facial expressions and emotion database, http://www.mmk.ei.tum.de/˜waf/fgnet/feedtum.html, technische universitat munchen," 2006. [Online]. Available: http://www.mmk.ei.tum.de/\~waf/fgnet/feedtum.html

[18] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, June 2010, pp. 94–101.

[19] K. L. Schmidt, Y. Liu, and J. F. Coh, "The role of structural facial asymmetry in in asymmetry peak facial expressions," *Literality*, vol. 11, no. 6, pp. 540–561, 2006.