EMOTION ANALYSIS ON HUMAN FACE THROUGH THERMAL IMAGING

Güler Püren, Dumlu Seda N. – Middle East Technical University
{e170986, e170987}@metu.edu.tr

Abstract

In this paper, emotion which can be categorized as valence and arousal, will be detected from the human face during a game playing session. Thermal camera measures the temperature changes according to different types of emotions such as valence or arousal. These temperature measurements are used in understand emotion changes of human. Analyzing human emotions using thermal infrared technology is useful in understanding people who has disabilities like autism or paralyzer. It is not only crucial for detecting disorders but it is also helpful for developing more complex robots that understands human emotions. Using thermal camera in analyzing human emotions through temperature changes in face is getting more popular since it is cheaper and more reachable than other methods like functional magnetic resonance imaging (fMRI). The hypothesis in this research lies on this mechanism; the researchers of this study wonder whether the blood perfusion on the face regions is correlated with the anxiety and stress levels of humans for the game technology or not. In this research, for fetaure extraction, co-occurence matrices of the region of interests (ROIs) in each frame are calculated. Energy, entropy, contrast, homogeneity, correlation features are calculated based on co-occurence matrix. Thus, the dataset is obtained. ROIs are the regions where the temparature change can be seen in the face. Then, principle component analysis (PCA) is applied for reducing dimensionality of fetaure set. Lastly, for discrimation of the frames where the player is excited (exciting frames) from the frames where the player is not excited (non-exciting frames), k-means clustering is applied to the dataset. According to the findings, the autors are able to discriminate the exciting frames with 75% ratio. However, 50% of the non-exciting frames are labeled as exciting frames.

Background Information

In the case of human computer interaction, face expression gains more importance for the computer's point of view since it can be easily captured. According to current literature, Pavlidis, Levine, and Baukol (2001) and Tsiamyrtzis et al. (2007) has shown that the blood perfusion in the orbital muscles is correlated with the anxiety and stress levels of humans.

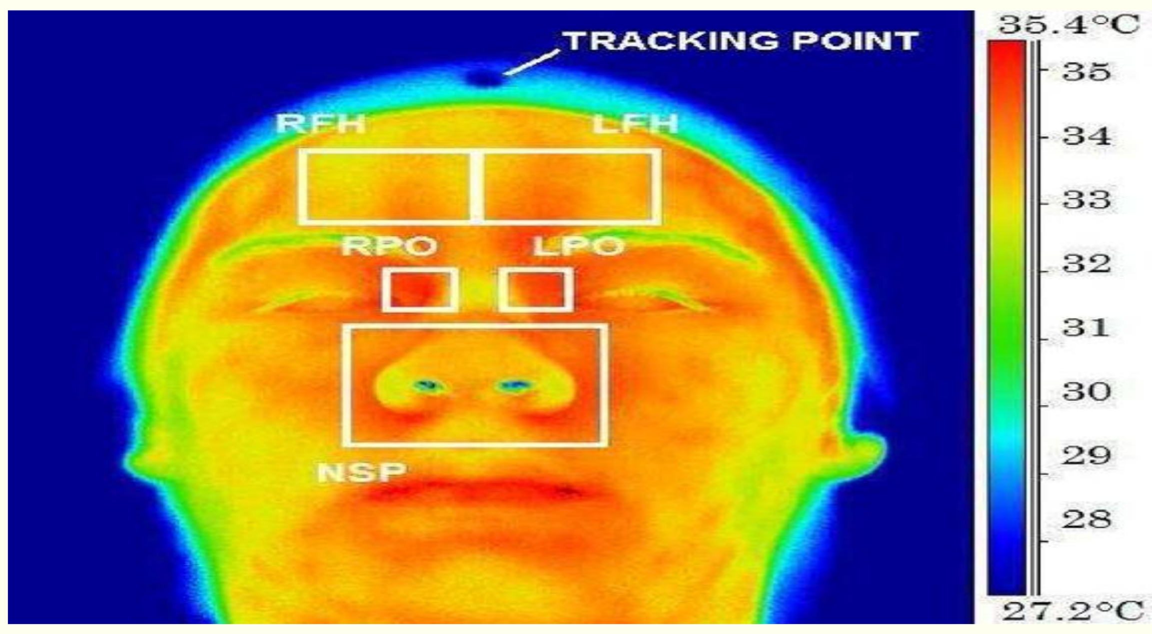

Figure 1 - Regions of interest used to calculated various statistical time series: left supraorbital (LFH), right supraorbital (RFH), left periorbital (LPO), right periorbital (RPO), and nasal (NSP). Adapted from [1]

In [1], they analyze the stress level of people during game-playing sessions in order to increase the entertainment level in games by designating proper difficulty levels for users. They measure stress level using thermal imaging of facial expressions. In their study, system measures player's stress and makes adjustment in game difficulty according to the stress level of the player. According to their results, their system is successful in understanding the stress level from the output of thermal camera and is able to give a pleasant game for the player by adjusting the game difficulty. Thus, their findings are a reference point for this research.

Experiment setup

For data collection, the subjects' faces are recorded with a thermal camera while they are playing a video game. The thermal camera is OPGAL EYE-R640 un-cooled infrared camera which captures 25 frames per second.

First the subject face is cooled. Then a neck rest is put on the neck of the subject for minimum head movement. Later, the game which is a war game is started and the subject face is recorded during game playing session. In the first experiment setup, the subject is recorded with only thermal camera, however due to tracking problem of face part for feature extraction, another experiment setup is created and this time, the subject face is also recorded with a visible camera.

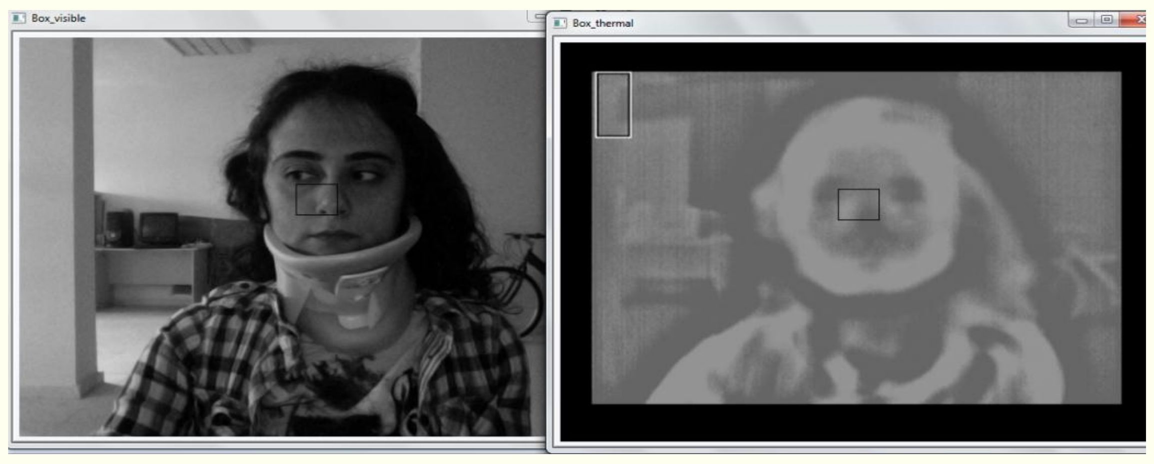

Figure 2 – The left picture shows the frame coming from visible camera and the right picture shows the frame coming from thermal camera. The blcak rectangle is the tracking window.

Methodology

Tracking of ROIs

For discriminating non-exciting and exciting frames, there should be some features related to those frames. Features are extracted from the ROIs in each frame. The ROIs are decided according to Figure 1. In each frame, the areas around noise and forehead are tracked and fetaure extraction methods are applied to those regions. At first, a simple tracker which is Meanshift algorithm is used. For each subject video, in the first frame, the search window location for ROI is selected manually. However, in the subsequent frames, the search window slides to different regions other than the selected region as seen in Figure 2. Due to bad results obtained from Meanshift, Camshift function of Opencv is applied. Although, the face is tracked successfully in visible videos, when the lighting change, it looses the face and because in the video game, there are constant flashing, the tracking operation is failed again. Lastly, the tracking operation is done manually. In the first frame the ROI is selected and if in the consequence frames the face change position, the program is stopped by the user, and ROI is selected again. The implementation is done with C++ by the help of Opencv library.

Feature Extraction

During tracking operation, co-occurence matrix of ROI is calculated and energy, entropy, contrast, homogeneity and correlation features are calculated based on this matrix for each frame. As a result, a dataset with 6 features and *N* data (*N* is the number of frames for a subject's video) is obtained. The implementation is done in C++. Since, a survey that asks the subjects in which part of the game they are excited does not exist, there is not an available validation whether the labeled frames are really exciting one or not. Thus, the authors make an assumption in this point which is since the game is a war game, while the game progresses, player comes accross with more combat and gunfight. So, the authors accepted that the first 70% of the frames are non-exciting ones and the last 30% of the frames are exciting ones. The visuallisation of dataset of one of the video sequences can be seen in Figure 3. For visualization, PCA is done on the dataset. As seen in Figure 3, exciting and non-exciting labeled frames are coloured differently.

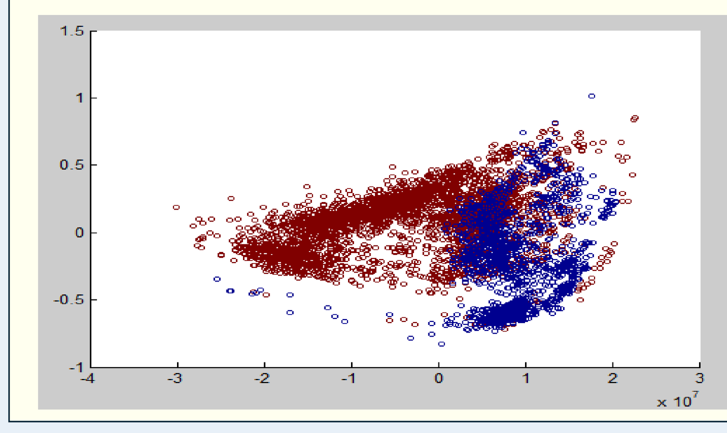

Figure 3 – Feature dataset of a video sequences. The %70 of the dataset is accepted as non-exciting and %30 is accepted as exciting. According to the figure. manual division does not give a very bad result

Methodology(cont.)

Discrimination

For discrimination of exciting and non-exciting features, k-means algorihm is used. The implementation is done with Matlab. One of the k-means result of the dataset can be seen in Figure 4. The dataset is clustered into two; exciting and non-exciting frames. As seen in the Figure 4, some manually labeled exciting and non exciting frames are clustered in the same cluster. In the figure, 890 of non-exciting frames are labeled as cluster 1 and 938 of non-exting frames are labeled as cluster 2 while 2260 of exciting frames are labeled as cluster 1 and 737 of them as cluster 2. Hence, since most of the non-exciting frames are clustered as 2 and most of the exciting frames are clustered as 1, cluster 1 blongs to non-exciting frames and cluster 2 blongs to exciting frames.

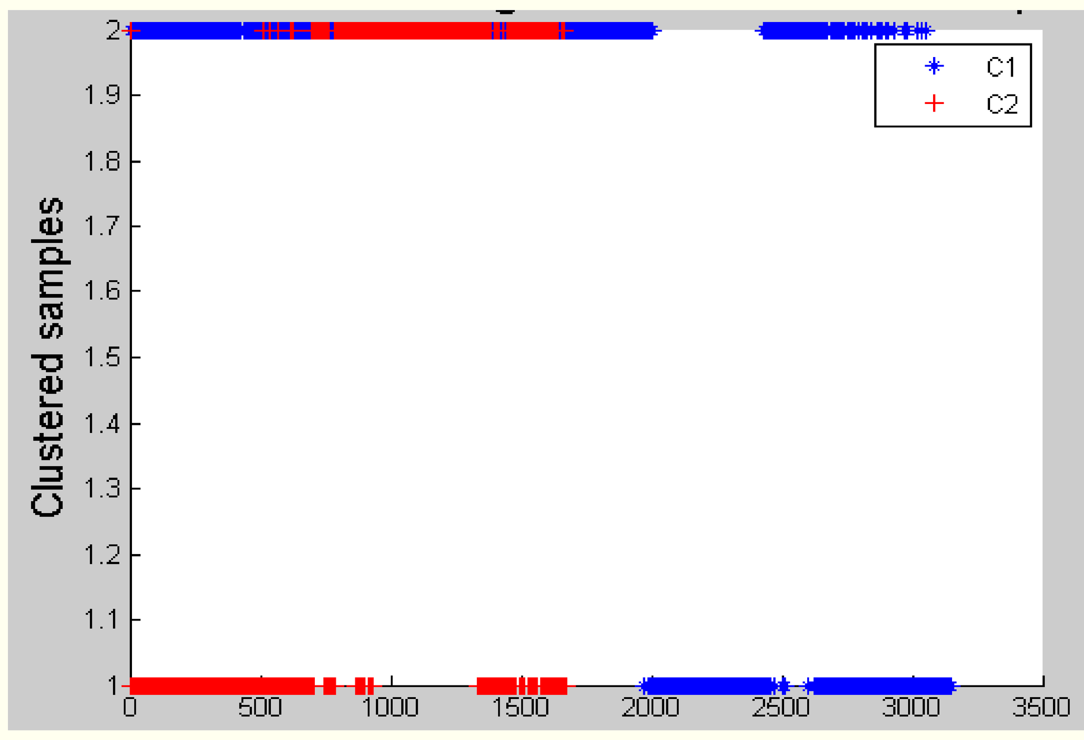

Figure 4 – one of the k-means result. The blue dots are the first frames(non-exciting). The red dots are the last frames(exciting). As we see from the figure there are some mislabeling like the blue dots which are non-exciting frames and red dots which are exciting frames are labeled in the same cluster. In th e below figures in

Result

In order to evaluate the cluster results, precision and recall are calculated for each video sequences. The number of correct labels are decided as described in Discrimination section.

		Correct classification	
		Exciting	Non-exciting
Obtained classification	Exciting	781	2595
	Non-Exciting	103	4405

Table 1- The result of one of the video sequences. Precision: 0,23, recall:0,88. Precision is very low because there are very misclassified item.

		Correct classification	
		Exciting	Non-exciting
Obtained classification	Exciting	2260	890
	Non-Exciting	737	938

Table 2- The result of one of the video sequences. Precision: 0,23, recall:0,88. Precision is very low because there are very misclassified item. Precision: 0,7, recall:0,75. In this classification, the correct classification is high relative to above one.

Conclusion

Although the authorrs do not have a proper validation whether the frames are non-exciting or exciting, with a rough splitting of the frames using k-means, an approximate idea about the frames are obtained. It is seen that although using primitive methods for classification, the most of the exciting frames are classified correctly. As a feature work, a new experiment setup for taking validation from subjects about in which part of the game they are excited. Also, a better tracking method should be applied.

Referance

[1] Yun, C.; Shastri, D.; Pavlidis, I.; Deng, Z.; , "O' Game, Can You Feel My Frustration?: Improving User's Gaming Experience via Stresscam," In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09). ACM, New York, NY, USA, pp. 2195-2204, 2009.

www.PosterPresentations.com