FACIAL EXPRESSIONS RECOGNITION BASED ON JOINT SHAPE-TEXTURE CUES IN VIDEO

LE ĐINH MINH ĐUC[*], NGUYEN TRUNG DUNG[**], LY QUOC NGOC[***]

ABSTRACT

This article proposed a facial expressions recognition system, which has a high recognition rate and can work in real-time, in frontal pose video. Our main contribution is that we present a new so-called ASL descriptor which is combined from three different basic features extraction method: AAM, SURF and LBP. Extensive experimental results demonstrate that our method achieve the performance higher than some state-of-art results on the Cohn-Kanade Database (CK).

Keywords: Active Appearance Models, Local Binary Pattern, Facial expression recognition, Real-time system, Speed-up Robust Features.

TÓM TẮT

Nhận dạng cảm xúc dựa vào đặc trưng dáng - vân trong video

Bài báo giới thiệu một hệ thống nhận diện cảm xúc khuôn mặt trong video với hướng nhìn trực diện, hệ thống này có độ chính xác cao và chạy trong thời gian thực. Đóng góp chính của bài báo là chúng tôi giới thiệu một đặc trưng mới ASL, ASL được kết hợp từ ba phương pháp rút trích đặc trưng cơ bản đó là: AAM, SURF and LBP. Kết quả thí nghiệm cho thấy rằng phương pháp này hiệu quả hơn những kết quả mới nhất của các hệ thống sử dụng bộ dữ liệu Cohn-Kanade Database (CK).

Từ khóa: Active Appearance Models (AAM), Local Binary Pattern (LBP), nhận diện cảm xúc khuôn mặt, hệ thống thời gian thực, Speed-up Robust Features (SURF).

1. Introduction

In recent years, facial expressions recognition is one of the most favorite topics in the field of computer vision. This contemporary area could be applied in many practical fields and might have a valuable meaning in human-machine interaction. This article would focus on the recognition of six basic emotions, namely: happiness, sadness, anger, surprise, disgust and fear. We will test the recognition rate of our system on Cohn-Kanade database (2000) [9] which is also very popular in scientific research community in the field of facial expressions recognition. When people express the emotion, there are several changes in their facial muscles and the shape of some facial components, such as: nose, mouth and eyes. Moreover, the wrinkles appear in

[*] B.S., in Computer Science at University of Science, Vietnam National University-Ho Chi Minh City.
[**] B.S., in Computer Science at University of Science, Vietnam national University-Ho Chi Minh City
[***] Dr., in Computer Science, University of Science,Vietnam National University-Ho Chi Minh City

some specific areas of human faces. We recognize the importance of both geometry feature and texture feature, so we proposes a new descriptor – ASL – by combine three basic feature extraction methods (AAM – SURF – LBP). This descriptor can also be inherited both local features (as texture features – SURF, LBP) and global feature (as geometric feature - AAM). Consequently, these texture features can solve the problem about scale variation while the geometric feature could give a hand to overcome the challenge of illumination variation and also increase the accuracy of our system. Finally, we use a statistical technique based on the accuracy of some selected image to decide a final result for all video.

2. Related works

Our discussion in this section will focus on feature extraction methods and experimental video databases.

Database:

In recent years, there are some facial emotion databases which were used very common in this field, such as: Cohn-Kanade database, MMI Facial Expression database [13] and JAFFE database [11], and each database has its own strengths. To be more detailed, CK database is a large database which contains over 480 image sequences and also has the challenge of illumination variation, age and skin color variation of the collaborators. All of emotion sequences in CK database are posed sequences, but non-posed sequences for several types of smiles have been added in the Extended Cohn-Kanade database [10]. Besides, MMI database is a very big database which has both posed sequences and non-posed sequences while JAFFE database is small and only helpful for laboratory experiments. This article uses CK database as the training and testing database because it has an AAM model and some challenges that I mention above.

Figure 1. *6 emotions in CK database*

Feature extraction:

According to Tian, Y., Kanede, T. and Cohn, J., [9] feature extraction might play the most important role in a facial expression recognition (FER) system. A feature is estimated as a good feature when the feature points in the same class have the minimum variation while the feature points from different classes have the maximum variation. In the past, the feature in FER could be divided into 2 main categories: geometry and texture.

Geometric features are more suitable for representing shape and location information of facial components. The scientists usually construct a model of landmarks (fiduciary points) in the human face and define an emotion by watching for the change of the positions of these points. Generally, Active Appearance Models (AAM) [6] and Active Shape Model (ASM) are very popular as geometric features.

On the other hand, texture features are more capable of capturing subtle appearance changes (e.g. wrinkles) of the face by digitizing these wrinkles when people express their emotion. Some common methods for extracting texture features are Gabor wavelet, PCA, LDA, SIFT… In particular, Gabor is also good but it has some drawbacks, such as: redundant; different performance contributions over different channels; expensive computation. In recent years, LBP [5][12] is very popular in this field because it has the advantage of tolerance against illumination changes, computational simplicity and also has a high performance in FER.

According to the above analysis, there are three options for feature extraction method:

(1) Extracting feature in the whole facial region.

(2) Extracting feature in the selected blocks in facial region.

(3) Extracting feature in the neighborhood of the fiduciary points.

The first option is not suitable for the system which contains variations in facial position and partial occlusions. In contrast, the second option could process the challenge of variations in facial pose, but it is very hard and flexible to choose the blocks to extract features. On the other hand, the final option might have a similar benefit as the second option, however it is better because it could have the clear areas to extract feature. As a result, we choose the third option as our approach.

3. Facial expressions recognition framework

Figure 2 shows the framework of our facial expressions recognition (FER) research and development system. It is consisted of four main steps: image pre-processing (face detection and face processing), feature extraction, feature selection and emotion classification. The input is a video and the output is a emotion of the person in the input video.

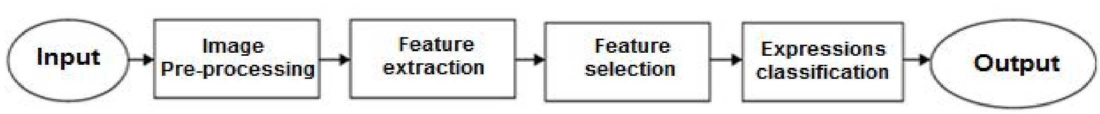

Figure 2. *Facial expressions recognition framework*

3.1. Face Pre-processing

Face detection and Facial region pre-processing play a very important role in our system because they could increase the performance of feature extraction.

3.1.1. Face detection

Face detection is used to locate the facial region in the selected frames of input video. Recent advances in face detection algorithms are mainly focus on Viola-Jones approach which utilized integral images to represent Haar-like features and adopted the Adaboost algorithm. After this step, we have a facial region which is cropped out to be processed in the following stage.

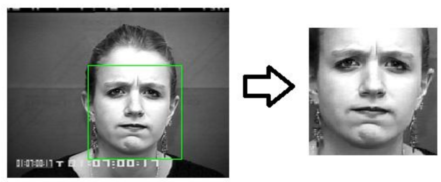

Figure 3. *Face detection Stage*

3.1.2. Facial region processing

The facial region might contain variations in illumination, size and position. By this step, we adjust the facial region of the selected frames into the same size. More importantly, we also use retinex algorithm to handle illumination variations and opening operator of mathematical morphology to thicken the edge.

Retinex algorithm [7] could balance the contrast and brightness of illumination of the facial region images. Particularly, this algorithm is imitated from the combination between retina and cortex which could help people to see clearly and distinguish color in the poor light conditions.

Multi - scale retinex (MSR):

$$R_{MSRi}(x,y) = \sum_{n=1}^{N} \omega_n \{\log I_i(x,y) - \log[F_n(x,y) * I_i(x,y)]\}$$

$I_i(x,y)$: the image distribution in the ith spectral band

$R_{MSRi}(x,y)$: Multi-scale retinex output

Gaussian function: $F(x,y) = Ke^{-(x^2+y^2)c^2}$, K determined by: $\iint F(x,y)dxdy = 1$, C is the Gaussian surround space constant.

N: number of scales.

ω_n: weight associated with nth scale.

Multi-scale retinex is better than single-scale retinex (N=1) in balance of dynamic compression and color rendition. As can be seen clearly from figure 4, after applying multi-scale retinex to facial region images, these images might be brighter if they were too dark, and vice versa.

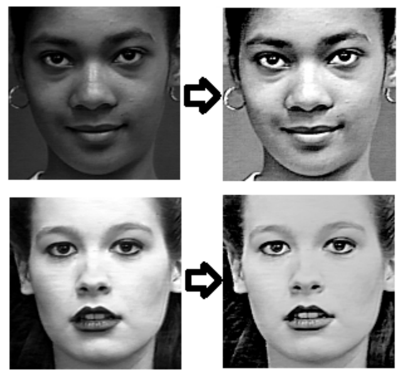

Figure 4. *The results of retinex algorithm*

Morphological opening [14]: opening is the dilation of the erosion of a gray-scale image f(x,y) by a structuring element b(x,y)

$$f \circ b = (f \ominus b) \oplus b$$

The erosion of f by a structuring element b at any location (x,y) is defined as the minimum value of the image in the region coincident with b when the origin of b is at (x,y). By contrast, the dilation of f by a structuring element b at any location (x,y) is defined as the maximum value of the image in the region coincident with $\hat{b} = b(-x, -y)$.

$$[f \ominus b](x,y) = \min_{(s,t)\in b}\{f(x+s, y+t)\}$$

$$[f \oplus b](x,y) = \max_{(s,t)\in b}\{f(x-s, y-t)\}$$

As a result, the eroded gray-scale image should be darker (bright feature is reduced and dark features are thickened) while the effects of dilation is opposite. Similar to erosion, opening might has a same effect on dark features (edges), however the edges of opening are smoother than the edges of erosion.

As can be seen clearly from the figure 5, the accuracy of AAM algorithm is better when we used retinex and morphological opening in the previous stage. This is mainly because the retinex algorithm could control the illumination of the facial region images while morphological opening would thicken the edge of these images. Moreover, AAM algorithm could distinguish the edge of facial components from other regions easily because color bins of grayscale image were reduced by applying morphological opening.

Figure 5. *Comparing the different between the accuracies of AAM*

Figure 5 shows the images 1a and 2a which are the results of AAM when we apply it into the original facial region images while images 1b and 2b are the results of AAM after the facial region images were processed by retinex and morphological opening.

3.2. Feature extraction

Feature extraction plays the most important step in facial expression recognition system. In this stage, we might extract the features which could help our system to distinguish 6 emotions. In this article, we combine 3 different feature extraction methods, namely: AAM, SURF and LBP to a new feature extraction method which is called ASL.

Emotions are expressed by the movement of the muscles in people faces. This movement also makes wrinkles and changes the shapes of the facial components, such as: nose, mouth, eyes and eyebrows. Firstly, we use AAM to define fiduciary points because the neighborhood of each point contains the most significant change. Then, we use SURF algorithm to recognize the direction of the change of the facial components and wrinkles. Finally, LBP algorithm is used to know the intensity of the change in shape of facial components and wrinkles.

Active Appearance Models [4] was introduced by Cootes, T. F., Edwards G. J. and Taylor C. J. (1998) which help us to generalize the shape and grayscale of a set of objects. AAM has two stages, namely: learning model and comparison model.

In the learning model stage, each image in CK database was marked by 61 fiduciary points, but we only use 51 fiduciary points to train the model. After that, we choose 17 fiduciary points from 51 points to apply the texture features in the neighborhood of these points. Importantly, the areas around these fiduciary points could cover all the important positions of a face to capture facial expressions.

Therefore, we could recognize the changes of texture features by using SURF and LBP into these areas. Although the number of fiduciary points is reduced, they are still consisted of the necessary points to capture the facial emotions. This fact also show the difference between features used in face recognition and facial expressions recognition.

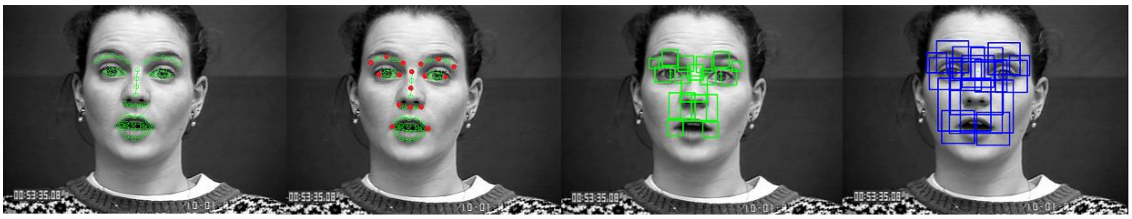

Figure 6. *An example of 17 fiduciary points*

Figure 6 shows the images which are the examples of using LBP and SURF around 17 fiduciary points.

Next, the process of comparison model depends on grayscale variation of feature extracted image and this is AAM model which is extracted from training dataset:

$$\delta \boldsymbol{I} = \boldsymbol{I}_i - \boldsymbol{I}_m$$

$\boldsymbol{I}_i$: The object that we want to find the value of it's parameter.

$\boldsymbol{I}_m$: The object represent for the training dataset.

$\boldsymbol{I}_i$ has a constant value in each frame and we find the parameter $\boldsymbol{I}_m$ to minimize the $\delta \boldsymbol{I}$ value in each frame**.** Therefore, we found $\boldsymbol{I}_m$ in the first frame before adjusting the range of parameter $\boldsymbol{I}_m$ in the following frames, so we could find it quickly and accurately.

In this article, *Speed-up Robust Features* [3] could recognize the direction of movement of the facial components and also texture in the neighborhood of each fiduciary point. SURF algorithm does not have the detection interest point step, but it still has the feature extraction stage. This mainly because these interest points were defined by AAM algorithm and they are 17 fiduciary points that I have mentioned above. These fiduciary points are invariant to rotation and illumination because 2 main reasons:

- The position of fiduciary points is unchanged;
- The structure is general for all faces.

We define a feature extraction area in the neighborhood of each fiduciary point. The size of this area is 20s, where s is a number that is chosen by experiment. Beside, this area is separated into 4 smaller areas and each new area is separated into 4 areas. After that, we have total 16 sub-areas. Next, we use 2s Haar-like feature to these areas to extract the final features. Finally, we receive a feature with 4 components for each area, so in total, we have 64 components of feature

The original Local Binary Pattern operator [12] which is quite simple but very effective was introduced by Ojala et al. (2002). In this article, LBP might show the intensity of the change of the facial components and texture in the neighborhood of each fiduciary point. The operator labels the pixels of an image by thresholding a 3x3 neighborhood of each pixel with the center value and considering the result as a binary number. This operator has two special characteristics which are invariant to illumination and could be used in a real-time system. Consequently, LBP is popular in computer vision field because it is very effective.

This is the process of feature extraction:

$$LBP_P = \sum_{p=0}^{P-1} s(g_p - g_c)2^p$$

Where g_c is the grayscale value of the center pixel.

- g_p is the grayscale value of neighbor pixels.
- $s(g_p - g_c)$ is a function which is defined as:

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

After labeling the image with the LBP operator, a histogram of labeled image $f_l(x,y)$ can be defined as:

Where l is the number of labels and function $I(A)$ is defined as:

$$I(A) = \begin{cases} 1, A\ is\ true \\ 0, A\ is\ false \end{cases}$$

The histograms of facial regions are also the features which were extracted by applying LBP operator.

Summary:

According to table 1, all of feature extraction methods have very short processing time, so they are very suitable for a real-time system. More importantly, we build a system which could make use of the strengths of each method. First of all, we use AAM to extract fiduciary points which is very important when people express emotion, because the neighborhood of each point changes significantly at this time. Then, we use SURF and LBP to recognize the change in these areas.

$$H_i = \sum_{x,y} I(f_l(x,y) = i), \qquad i = 0,1,\dots,l-1$$

Table 1. *Characteristics of three selected feature extraction methods*

References	Feature extraction method	Role	Strong points	Adaptation with this system
Cootes, T.F. , Edwards, G.J. và Taylor, C.J., 1998 [6]	Active Appearance Models (AAM)	Extract 17 fiduciary points. Guarantee for global based of our system.	Solve the problem about posture variation of collaborators. Need fewer fiduciary points than ASM algorithm.	Make use of information from the recent frame to find parameter vector for the following frame.
Bay, H., Ess, A., Tuytelaars, T. and Van Gol, L., 2008 [3]	Speed-Up Robust Features (SURF)	Recognize the information of grayscale variation in the neighborhood of each fiduciary point. Guarantee for local based of our system.	Feature Vector of SURF is shorter than Feature Vector of SIFT, but it also contains enough information. Time for processing of SURF is shorter than SIFT	Recognize the information of grayscale variation of wrinkles which appear when people express emotion.
Ojala, T., Pietikainen, M., Maenpaa, T., 2002 [12]	Local Binary Pattern (LBP)	Recognize the information of grayscale distribution in the neighborhood of each fiduciary point. Guarantee for local based of our system.	Time for processing is very fast. Invariant to illumination.	Recognize the information of grayscale distribution in the neighborhood of wrinkles.

The ASL descriptor achieves some properties such as:

- Invariant to illumination;
- Invariant to scale;
- High recognition rate;
- Real-time descriptor.

3.3. Feature selection

In this stage, we use *Principal Component Analysis (*PCA). PCA is a way of finding out which features are important for best describing the variance in a database or in a feature vector. It is most often used for reducing the dimensionalities of a feature vector so that it becomes more practical to apply machine learning where the original feature vector is inherently high dimensional. In this system, the threshold which determines the dimension of the feature vectors after applying PCA is about 0.98, so the dimensionalities of the original feature vector reduce from 5440 to only 369. We use PCA because some following benefits:

(1) PCA could reduce the dimensionalities of a feature vector;

(2) The new space is also similar to the original space;

(3) In the new space, the secret connection is found.

3.4. Expressions Classification

Support Vector Machines (SVM) [16] is a very common and helpful algorithm for FER problem. In my system, we use SVM to classify the emotion in the static image. SVM try to find a classified plane which could maximum the distance between two margins. The linear classifier predicts a new sample by this formula:

$$h(x) = sgn(\langle x \cdot w \rangle + b)$$

However, the data distribution in practice is very complex and we can't use a linear plane to classify emotion. The solution in this situation is that we map the data points to a new space which has more dimensions and fine a plane which could divide the linear data. The mapping of the data points are represented by the following formula:

$$K(x, z) = \langle \Phi\langle x \rangle \cdot \Phi\langle z \rangle \rangle$$

$K(x, z)$ is kernel function. In our study, we use the kernel function: radial basis function. We can rewrite the defined function of the classified plane as follow:

$$h(x) = sgn\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i \cdot x) + b\right)$$

Where α_i is the Lagrange coefficient.

SVM makes binary decisions, so we use one-vs-all method to classify 6 emotions in this stage. We train 5 binary classifiers from 6 classes of emotions to discriminate one expression from all others, and output is the class with the largest output of binary classification. For example, the first classifier is happiness-vs-5 other emotions while the second classifier is sadness-vs-4 other emotions…

Particularly, we used radial basic kernel with the kernel's parameter γ and soft margin parameter C:

$$K(x,z) = e^{-\gamma\|x-z\|^2}$$

We recommend a "grid-search" on C and γ using cross-validation. Various pairs of (C,γ) values are tried and the one with the best cross-validation accuracy is picked. Consequently, we found the best $(C = 300, \gamma = 0.02)$ for this system on $C \in \{100, 150, 200, \dots, 500\}$ and $\gamma \in \{0.01, 0.02, \dots, 0.05\}$.

4. Experimental Results

In this article, we evaluate the performance of our system for both static image and video in the CK database. The experiment result of ASL facial expression recognition system for static image is showed in the table 2. As can be seen clearly, 4 emotions, namely: happiness, anger, surprise and disgust have the highest accuracy. The average accuracy of our system is around 90.96%. Moreover, we also test the performance of the systems which use AAM and SURF (Table 3) or AAM and LBP (Table 4) in the feature extraction stage. Consequently, the result of our system is better than other systems which use only two feature extraction method.

Experiments on static image:

Table 2. *The accuracy of our system (AAM, SURF and LBP)*

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	**93.33%**	4.44%	0.00%	0.00%	2.22%	0.00%
Disgust	0.85%	**97.46%**	1.69%	0.00%	0.00%	0.00%
Fear	2.00%	0.00%	**88.00%**	2.00%	4.00%	4.00%
Happiness	0.00%	0.00%	1.45%	**98.55%**	0.00%	0.00%
Sadness	17.86%	0.00%	7.14%	0.00%	**71.43%**	3.57%
Surprise	0.00%	0.00%	1.81%	1.20%	0.00%	**96.99%**

Table 3. *The accuracy of the system which uses AAM and SURF in the feature extraction stage*

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	**88.89%**	6.67%	0.00%	0.00%	4.44%	0.00%
Disgust	4.24%	**94.92%**	0.85%	0.00%	0.00%	0.00%
Fear	2.00%	0.00%	**88.00%**	2.00%	4.00%	4.00%
Happiness	0.00%	0.00%	1.45%	**98.55%**	0.00%	0.00%
Sadness	21.43%	1.79%	5.36%	0.00%	**67.68%**	3.57%
Surprise	0.60%	0.00%	0.60%	1.20%	1.20%	**96.39%**

Table 4. *The accuracy of the system which uses AAM and LBP in the feature extraction stage*

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	**93.33%**	2.22%	1.11%	0.00%	2.22%	1.11%
Disgust	6.78%	**88.14%**	2.54%	1.69%	0.85%	0.00%
Fear	8.00%	4.00%	**64.00%**	4.00%	12.00%	8.00%
Happiness	1.45%	0.00%	2.17%	**94.93%**	0.00%	1.45%
Sadness	19.64%	0.00%	17.86%	0.00%	**60.71%**	1.79%
Surprise	1.20%	0.00%	4.82%	0.00%	0.00%	**93.39%**

Table 5. *Comparing between our system and related systems for static image in recent years*

References	**Feature extraction method**	**Database**	**Accuracy**
Caifeng Shan et al. (2009) [5]	Boosted-LBP	CK	Almost 88.67%
Cohn and Kanade (2010) [10]	AAM (SPTS + CAPP)	CK	Almost 83.15%.
Abdat et al. (2011) [1]	Change of 21 distances between the key points which were extracted by FACS method	CK	About 95%
Songfan Yang and Bir Bhanu (2012) [15]	Emotion avatar image (EAI use SIFT flow) + LBP or EAI + local phase quantization (LPQ)	CK+	EAI+LPQ: 82.6%.
Our system for static image	ASL (AAM+SURF+LBP)	CK	About 90.96%

As can be seen clearly from the table 5, our system could overcome most of previous systems for static image and it also is a real-time system.

Experiments on video: We use the statistical technique on the second half of image sequence to receive the final result for video. Table 6 shows the result of our system for video.

Table 6. *The accuracy of the system for video*

	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Anger	**100.0%**	0.00%	0.00%	0.00%	0.00%	0.00%
Disgust	3.39%	**96.61%**	0.00%	0.00%	0.00%	0.00%
Fear	0.00%	4.00%	**96.00%**	0.00%	0.00%	0.00%
Happiness	0.00%	0.00%	1.45%	**98.55%**	0.00%	0.00%
Sadness	0.00%	0.00%	0.00%	3.57%	**96.43%**	0.00%
Surprise	0.00%	0.00%	0.00%	0.00%	1.20%	**98.80%**

The face expressions recognition on video achieves the accuracy higher than on static image because the final results are voted not only from one image but also from many frames on video. The accuracy of our system on video achieves 97.73% and table 7 shows the comparison between our system and other systems for video.

Table 7. *The accuracy of the system for video*

References	Feature extraction method	Accuracy
Yeasin et al. (2004) [17]	Hidden Markov Models (HMMs)	91.2%
Aleksic et al. (2006) [2]	Multi-stream HMMs	93.66%
Buenaposada et al. (2008) [4]	Bayesian	89%
Guoying Zhao et al. (2009) [8]	Multi-resolution Spatiotemporal Space	93.85%
Our system for video	ASL	97.73%

Finally, our system could handle 30 frames in the average time of 1.6 seconds.

5. Conclusion and future work

This article proposes a facial expression recognition system for frontal pose in video. The most important contribution of our system is a new feature extraction method that could not only have a very high recognition rate but could also work in real-time. Besides, we introduce 17 selected fiduciary points from 51 fiduciary points in AAM to develop a real-time system. Our system is a real-time system which achieves the performance higher than some state-of-art results.

In future works, we would like to develop our system on RGB-D video. Moreover, we will use other classification methods to demonstrate the performance and value of our system.

TÀI LIỆU THAM KHẢO

1. Abdat, F., Maaoui C. and Prusk, A. (2011), *"Human-computer interaction using emotion recognition from facial expression"*, UKSim 5th European Symposium on Computer Modeling and Simulation, pp. 196-201.
2. Aleksic, S.P., Katsaggelos, K.A., (2006*),* *"Automatic facial expression recognition using facial animation parameters and multi-stream HMMS"*. IEEE Trans. Information Forensics and Security 1 (1), pp. 3–11.
3. Bay, H., Ess, A., Tuytelaars, T. and Van Gol, L. (2006), Surf: *"Speed up robust features"*. European Conference on Computer Vision, 1:404-417.
4. Buenaposada, J.M., Muñoz, E., Baumela, L. (2008), *"Recognising facial expressions in video sequences"*, Pattern Anal. Appl. 11, pp. 101–116.
5. Caifeng Shan, Shaogang Gong, Peter W. McOwan (2009), *"Facial expression recognition based on Local Binary Patterns: A comprehensive study"*, Image and Vision Computing, Vol. 27, pp. 803-816.
6. Cootes, T. F., Edwards, G. J. and Taylor. C. J. (1998), *"Active appearance models"*, In Proc. European Conf. on Computer Vision, Vol. 2, pp. 484-498. Springer.
7. Edwin H. Land (1997), "The Retinex Theory of Color Vision", Scientific American, Vol. 237, No. 6, December, pp. 108-128.
8. Guoying Zhao, Matti Pietikäinen (2009), *"Boosted multi-resolution spatiotemporal descriptors for facial expression recognition"*, Pattern Recognition Letters 30, pp. 1117–1127.
9. Kanade, T., Cohn, J. F., & Tian, Y. (2000), *"Comprehensive database for facial expression analysis"*, Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, pp. 46-53.
10. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010*),* *"The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression"*, Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, pp. 94-101.
11. Michael J. Lyons, Shigeru Akamatsu (1998), *"Coding facial expressions with Gabor Wavelets"*, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.

12. Ojala, T., Pietikainen, M., Maenpaa, T. (2002), *"Multiresolution gray-scale and rotation invariant texture classification with local binary patterns"*, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7), pp. 971-987.
13. Pantic, M., Valstar, M.F., Rademaker, R. and Maat, L. (2005), *"Web-Based Database for Facial Expression Analysis,"* Proc. 13th ACM Int'l Conf. Multimedia (Multimedia '05), pp. 317-321.
14. Rafael C. Gonzalez and Richard E. Woods (2008), *"Digital Image Processing – Third Edition"*, International Edition, August, pp. 627-688.
15. Songfan Yang and Bir Bhanu (2012), *"Understanding Discrete Facial Expressions in Video Using an Emotion Avatar Image"*, IEEE Transactions on systems, man, and cybernetics — PART B: Cybernetics, Vol. 42, No. 4, August.
16. Vapnik, V. and Cortes, C. (1995), *"Support-Vector Networks"* Machine Learning Vol. 20, pp. 273-297.
17. Yeasin, M., Bullot, B., Sharma, R., (2004), *"From facial expression to level of interest: A spatio-temporal approach"*. In: Proc. Conf. Computer Vision and Pattern Recognition, pp. 922–927.

(*Received: 14/9/2013; Revised: 21/11/2013; Accepted: 13/12/2013)*