Facial Expression Recognition of Animated Human Characters

Guoxu Liu[1,2], Hui Jin[1], Hailong Jiang[1] and Jae Ho Kim*[1]
[1]Dept. of Electronics Engineering, Pusan National University, Busan, Rep. of Korea
[2]Computer Software Institute, Weifang University of Science & Technology, Shouguang, China
201693257lgx@pusan.ac.kr; lay1007@pusan.ac.kr; pandalgx@gmail.com; jhkim@pusan.ac.kr

ABSTRACT
Recognition of animated facial expression is an important and a fundamental task for automatically implementing analysis and evaluation of animation movie products. In this work, an improved algorithm was proposed for recognizing facial expression of animated human characters based on the Local Binary Pattern (LBP) and Support Vector Machine (SVM). The proposed method was tested on an animated facial expression database which contained human characters from top rated animations. The experimental results showed that the recognition accuracy of facial expression was significantly improved compared with other methods.

CCS Concepts
• Computing methodologies → Appearance and texture representations • Computing methodologies → Object recognition • Computing methodologies → Support vector machines

Keywords
Facial expression recognition; animation movie analysis; multiple range LBP

1. INTRODUCTION
Nowadays, animation is developing rapidly, and it is a big demand of corporations to produce animations of high quality for the audience. The most important content of an animation movie is the plot of the story which is the primary impression for audience to enjoy [1]. It is necessary to create an analysis system to evaluate the quality of animation movies or episodes. However, thousands of animation products are produced in each year. It is almost impossible to evaluate every animation movie or episode of animation series by manual inspection. Therefore, it is necessary to build an automated system for animation evaluation. Fortunately, animation products are digitalized, indicating that the automatic evaluation system is possible by applying the state-of-the-art image analyzing techniques and machine learning methods.

ICMLC 2020, February 15–17, 2020, Shenzhen, China

ACM ISBN 978-1-4503-7642-6/20/02…$15.00
DOI: https://doi.org/10.1145/3383972.3384026

Actions and emotions of the animated characters are directly related to the story. Therefore, facial expression, which represents the emotional variation of the character, is an important cue to describe the scene or the story in the animation products. Facial expression recognition of animated faces is necessary to implement automatic animation analysis and evaluation systems. However, recognition of animated characters has not been widely studied.

In this study, we aim to provide a technical solution for recognizing facial expressions of animated human characters. This study remedied the shortcoming of the traditional LBP descriptor on the nonuniform texture, by proposing the multiple-range LBP and multiple-region LBP. Besides, an SVM classifier was adopted for classification. Six frequently appearing facial expressions, which included: anger, disgust, fear, happy, sad and surprise were selected as the recognition target in this work. Human characters from high rated animation movies produced by corporations in USA were selected in this work for experiments.

The remainder of this paper is organized as follows. Section 2 presents the related work. Section 3 describes the proposed facial emotion recognition method. Section 4 reports and discusses the experimental results. And section 5 draws conclusions from this paper.

2. RELATED WORK
Animated human characters are created based on real human and present the same structure with human faces. Although the details of facial mechanisms of the animated characters are rather different from that of real humans in order to enlarge the emotional expressions and emphasize the circumstances in animations, both the faces of the animated characters and real human could present the similar visual features for the audience to understand their emotions. Therefore, developing a recognition algorithm for animated characters should refer to the previous works on human facial expression recognition [2].

A typical scheme used for human facial recognition is shown in Figure 1. There are two phases for constructing a facial recognition system. The first step is to extract the features of the face and the second step is to train the classifier with the features. Other steps such as noise removal and image enhancement can also be performed as a preprocessing [3].

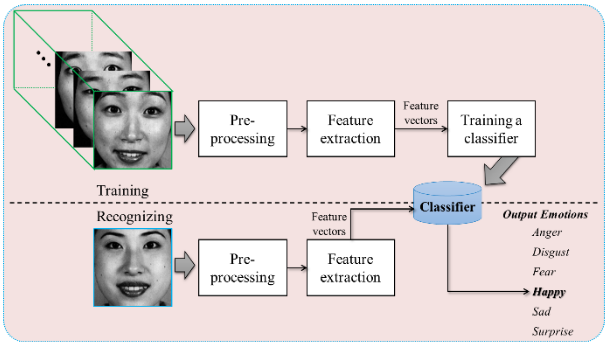


Figure 1. A typical facial expression recognition system.

Many studies have been researched on the feature extraction for vision tasks [4-9]. In general, there are two kinds of features for facial expression recognition, i.e. geometric feature and appearance feature. The shapes and geometric locations of the facial components are the main information in the geometric features. The graph-based method is one of the geometric feature extraction methods and has been widely studied [5-7]. Valstar proposed a facial action coding based method for facial recognition and achieved an excellent accuracy [8,9]. In the later study, deformable models were utilized to produce accurate descriptions of facial components [10,11]. A well-known geometric deformable model, Active Shape Models (ASM), was proposed by Cootes et al. [12]. The ASM showed outstanding accuracy and stability for extracting facial features [13]. However, the geometric feature-based methods usually require accurate and reliable facial feature for detection, which is difficult to accommodate in many situations.

On the other side, appearance-based features were proposed to extract entire appearance features from the whole face images or create local feature from specific facial component (e.g., eyes, brow or mouth) to describe the appearance changes in each type of facial expression. Some methods extracted the features from the entire faces for recognition, such as "eigen faces" and "fisher faces" [14,15]. The linear discriminant analysis is another widely applied global descriptor for facial recognition [16]. Some other researchers used local descriptors for facial expression recognition [2]. The accuracy of facial recognition has improved significantly when local facial descriptors were incorporated, comparing with that of only using global features [17,18]. Many effective local feature descriptors have been applied to recognize facial expression including Scale Invariant Feature Transformation (SIFT) [19], Gabor features [20], elastic bunch graph matching [21], and LBP [22].

The LBP was originally designed for local texture description, which was afterwards applied to facial recognition. Since LBP achieved outstanding recognition performance compared to other local feature descriptors, the LBP based facial recognition methods have been studied extensively to improve the recognition accuracy [23,24]. Therefore, in this work, we studied the facial expression recognition of animated faces based on the LBP descriptor and proposed an improved recognition scheme.

Another essential part in the facial expression recognition system is the classification algorithm. Different techniques have been studied for the classification of facial expressions or other tasks, such as Bayesian Network (BN) [25], Artificial Neural Network (ANN) [26,27] and SVM [28-30]. The SVM is a classical machine learning algorithm based on the statistical learning theory. According to the principle of minimizing Vapnik structure risk [31], the SVM makes a tradeoff in function complexity and sample complexity, improving the generalization ability of learning and achieving excellent classification performance. The SVM has proved its excellent classification accuracy and stability in many related studies [32,33]. Thus, in this work, we utilized the SVM as the facial expression classifier.

3. PROPOSED METHOD

The original LBP calculates only local variations on gray values. Therefore, the input images are usually required to be divided into numbers of small sub-images to obtain the local features over the whole images. And these features are combined as the feature vector to describe the feature of whole images. This feature extraction scheme requires that the distribution of texture should be uniform or less of changes. For example, facial components should be presented in the proper positions without much offset on locations. For this reason, the LBP descriptor demonstrates an excellent performance on human facial expression recognition since the facial components on human faces (e.g., eyes, brows) do not have significant variations with respective to positions and shapes. Especially, human faces in the public database are posed in the same angle and light conditions which make the recognition task much simple. On the contrary, animated faces are rather challenging which could not satisfy the requirement of the local feature descriptors above due to the exaggerated facial expression.

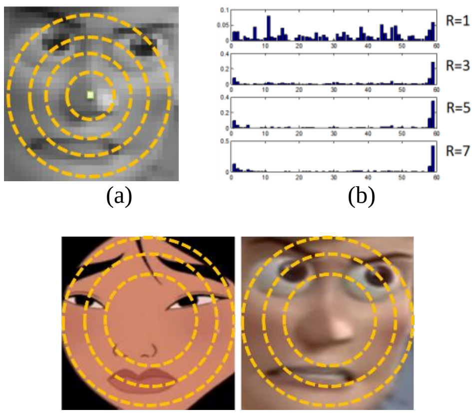
3.1 Proposed LBP Based Animated Facial Feature

Although significant variances of facial components are presented on animated faces, the overall patterns of each type of emotions are comparable. Thus, the global facial feature of the whole image should also be considered in the recognition of animated faces. Instead of the original LBP, we proposed a multiple-range LBP method. This method calculates LBP with multiple radiuses and integrates the LBP with multiple regions. These multiple-range LBPs could compose the overall features of faces.

3.1.1 Multiple-radius LBP

The conventional LBP only calculate the intensity variations on its 3×3 neighbors for each pixel within a sub-window. We propose to extend the neighborhood for LBP by increasing the radius. We calculated the LBPs with various radiuses to describe the overall pattern of emotions. The multiple-radius LBP was calculated from the center point of the facial images to obtain the global feature description (Figure 2(a)). The original LBP could only represent local variation in intensity around a pixel that is impossible to describe the global facial pattern and the relations among facial components. With the proposed multiple-radius LBP, the LBP feature is calculated by various radius which is illustrated as the yellow circles in Figure 2(a). Each circle could reflect the facial pattern at different ranges. For example, the inner circle might describe only the feature of nose while the medium two circles contain eyes and mouth, and the feature of the brows could be additionally covered by the outer circle. Therefore, more than one facial component could be described by a circle that constructs the relation among these facial components. Accordingly, the overall facial pattern could be represented by combining all these multiple-radius LBPs. Numerical examples of multiple-radius LBP are presented in Figure 2(b) that histograms of LBP with multiple-radius present face features at different levels.

The multiple-radius LBP is possible to describe the flexible facial components in various locations. As shown in Figure 2(c), although the eyes are presented in very different locations in the female and male faces, they can be covered by the medium range circles. The female face shows sad emotion and the male face is surprise. Their brows and eyes present different shapes and texture that they could be modeled by the proposed multiple-radius LBP. The facial components will present different LBP patterns on each level of radius, such as female showing upward brow while male's brow is flat. In order to cover the whole image by the radius we transform the original image into a square image (e.g. 160 × 160 pixels).



(c)

Figure 2. Facial feature calculation with the proposed multiple radius LBP and the corresponding histograms.

3.1.2 Multiple-overlapped-region LBP

In addition to the multiple-radius LBP, the LBP from multiple overlapped regions are proposed to examine the features of face expression variations as a complementary to multiple-radius LBP. As illustrated in Figure 3, the facial image is initially equally divided into four quarters. The top two quarters extract the feature of two eyes and brows on the left and right side. The bottom two quarters mainly describes the mouth in combination. And the central region is utilized to obtain the feature of nose with a size equal to a quarter. These multiple regions correspond to the main facial components. Since facial components show significant variations, large-sized (a quarter) region is assigned to ensure the feature of each facial component could be captured in the regions.

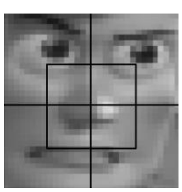


Figure 3. Multiple regions for obtaining facial components feature by LBP.

4. EXPERIMENTAL RESULTS

4.1 Dataset

In order to study the facial recognition for practical applications and to reveal the common pattern of animated characters, we selected the top-rated movies from 1990 to 2016 covering various topics including historical story, fairytale, children's story, and science fiction, etc. The selected movies are given in Table 1. The criteria of selecting animation movies are: (1) containing animated human characters, (2) box office > 300 million dollars, and (3) IMDB rating ≥ 7.5.

Table 1. Selected animation movies for facial recognition of animated characters.

Movie title	Premiere	Box Office ($)	IMDB rating
Beauty and the Beast	1991-11-22	424,967,620	8
Aladdin	1991-11-25	501,900,000	8
Toy Story	1995-11-22	361,958,736	8.3
Mulan	1998-06-19	303,500,000	7.5
Ratatouille	2007-10-19	623,722,818	8
Up	2009-05-13	731,342,744	8.3
How to Train Your Dragon	2010-05-14	494,878,759	8.2
Tangled	2010-11-24	591,794,936	7.8
Wreck-It Ralph	2012-11-01	471,222,889	7.8
Frozen	2013-11-27	1,276,480,335	7.6
Big Hero	2014-06-01	657,818,612	7.9

The facial images of the characters were manually cropped from the movie frame. Frontal faces were preferred when collecting faces from movies. However, frontal faces were not frequently available in the movies, so some slightly rotated or side view faces were also collected into the facial expression database.

The facial image dataset contains six basic kinds of emotions, i.e. anger, disgust, fear, happy, sad and surprise. For each type of facial expression, 74 facial images were selected. The animated facial expression dataset in this work contains 444 facial images in total. Some examples are shown in Figure 4.

Figure 4. Some examples of collected facial images.

4.2 Comparison with Other Methods

To validate the efficiency of the proposed algorithm, another method [2] was compared. Moreover, one of the popular deep learning methods – VGG16 [34] – was also applied to evaluate its performance. The recognition results of the methods are given in Table 2. It showed that the recognition accuracy was significantly improved using the proposed method.

As presented in Figure 5, some of the difficult faces missed while using original LBP were recognized by our proposed method. The facial components on these faces showed irregular locations and some of them were with side views. For example, the first three faces showed small eyes, which were in the center of the faces while the last female face was with rather big eyes located on the side of the face. These faces presented significant variations which were incorrectly classified with the LBP in [2] but were correctly identified by our proposed methods.

Besides, the VGG16 showed poor results. This is mainly due to the small size of the dataset, which results in an underfit on the network. So, for a specific task especially in possession of small size of data, traditional machine learning methods are preferred than deep learning methods.

However, since the facial expressions in animations are rather complicated, the recognition accuracy is not comparable to human database. Too many side-view faces were difficult to identify. When beard, glasses and hair present on the faces, the recognition accuracy would be decreased.

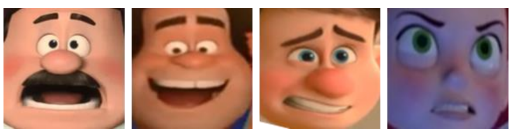


Figure 5. Some examples of improved facial recognition by the proposed method.

Table 2: Facial expression recognition of animated faces using traditional LBP and proposed method.

Facial Emotion	Traditional LBP [2] (%)	VGG16 [34] (%)	Proposed Method (%)
Anger	100	75.67	97.3
Disgust	24.32	21.62	83.8
Fear	18.92	32.43	81.1
Happy	70.27	64.86	86.5
Sad	40.54	64.86	89.2
Surprise	48.65	59.45	91.9
Average	50.45	53.15	88.3

5. CONCLUSIONS AND FUTURE WORKS

In this paper, the recognition of animated human character's facial expression was studied based on the LBP descriptor and SVM. To address the limited representation ability of the traditional LBP, a multiple-range LBP descriptor was proposed to obtain global facial features for recognition of the animated human facial expression. The multiple-range LBP was calculated with various radius and multiple regions were assigned to acquire additional facial features. The proposed facial recognition scheme was tested on an animated facial expression database which contained human characters from top rated animations. The experimental results showed that the overall recognition accuracy of the facial expression was significantly improved compared to the traditional LBP. Moreover, a comparison experiment with the deep learning method was also set up, and the results showed that for specific small size of data, the traditional learning method could get better results, and thus can be fast deployed for application.

However, the accuracy of the proposed method is still unsatisfactory compared with that of human facial expression recognition and needs to be further improved. In the future, the transfer learning method will be adopted with an extension of the dataset.

6. ACKNOWLEDGMENTS

This work was supported by BK21PLUS of Pusan National University.

7. REFERENCES

[1] Sung, B., Kim, M. and Kim, J., 2013. Classification model of the 17 functional program narrative of animation. *Archives of Design Research*, 26(1).

[2] Shan, C., Gong, S. and McOwan, P.W., 2009. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and vision Computing*, *27*(6), pp.803-816.

[3] Bow, S.T. ed., 2002. *Pattern recognition and image preprocessing*. CRC press.

[4] Ojala, T., Pietikäinen, M. and Mäenpää, T., 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7), pp.971-987.

[5] Kakumanu, P. and Bourbakis, N., 2006, November. A local-global graph approach for facial expression recognition. In *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)* (pp. 685-692). IEEE.

[6] N. G. Bourbakis, P. Kakumanu, 2008. *Applied Pattern Recognition*. Springer-Verlag, New York.

[7] Bourbakis, N., Esposito, A. and Kavraki, D., 2011. Extracting and associating meta-features for understanding people's emotional behaviour: Face and speech. *Cognitive Computation*, *3*(3), pp.436-448.

[8] Valstar, M.F., Patras, I. and Pantic, M., 2005, September. Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops* (pp. 76-76). IEEE.

[9] Valstar, M. and Pantic, M., 2006, June. Fully automatic facial action unit detection and temporal analysis. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)* (pp. 149-149). IEEE.

[10] Fasel, B. and Luettin, J., 2003. Automatic facial expression analysis: a survey. *Pattern recognition*, *36*(1), pp.259-275.

[11] Kotsia, I. and Pitas, I., 2006. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE transactions on image processing*, *16*(1), pp.172-187.

[12] Cootes, T.F., Taylor, C.J., Cooper, D.H. and Graham, J., 1995. Active shape models-their training and application. *Computer vision and image understanding*, *61*(1), pp.38-59.

[13] Milborrow, S. and Nicolls, F., 2008, October. Locating facial features with an extended active shape model. In *European conference on computer vision* (pp. 504-513). Springer, Berlin, Heidelberg.

[14] Turk, M. and Pentland, A., 1991. Eigenfaces for recognition. *Journal of cognitive neuroscience*, *3*(1), pp.71-86.

[15] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: Recognition using class specific

linear projection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7), pp.711-720.

[16] Etemad, K. and Chellappa, R., 1997. Discriminant analysis for recognition of human face images. *Josa a*, *14*(8), pp.1724-1733.

[17] Heisele, B., Ho, P., Wu, J. and Poggio, T., 2003. Face recognition: component-based versus global approaches. *Computer vision and image understanding*, *91*(1-2), pp.6-21.

[18] Zou, J., Ji, Q. and Nagy, G., 2007. A comparative study of local matching approach for face recognition. *IEEE Transactions on image processing*, *16*(10), pp.2617-2628.

[19] Li, Z., Imai, J.I. and Kaneko, M., 2010. Facial Expression Recognition Using Facial-component-based Bag of Words and PHOG Descriptors. *Information and Media Technologies*, *5*(3), pp.1003-1009.

[20] Gabor, D., 1946. Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, *93*(26), pp.429-441.

[21] Wiskott, L., Fellous, J.M., Krüger, N. and von der Malsburg, C., 1997. Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), pp.775-779.

[22] Feng, X., Hadid, A. and Pietikäinen, M., 2004, September. A coarse-to-fine classification scheme for facial expression recognition. In *International Conference Image Analysis and Recognition* (pp. 668-675). Springer, Berlin, Heidelberg.

[23] Huang, D., Shan, C., Ardabilian, M., Wang, Y. and Chen, L., 2011. Local binary patterns and its application to facial image analysis: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(6), pp.765-781.

[24] Tang, H., Yin, B., Sun, Y. and Hu, Y., 2013. 3D face recognition using local binary patterns. *Signal Processing*, *93*(8), pp.2190-2198.

[25] Cohen, I., Sebe, N., Garg, A., Chen, L.S. and Huang, T.S., 2003. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and image understanding*, *91*(1-2), pp.160-187.

[26] Steur, N.A. and Mueller, C., 2019. Classification of Viral Hemorrhagic Fever Focusing Ebola and Lassa Fever Using Neural Networks. *International Journal of Machine Learning and Computing*, *9*(3).

[27] Kharat, G.U. and Dudul, S.V., 2009. Emotion recognition from facial expression using neural networks. In *Human-Computer Systems Interaction* (pp. 207-219). Springer, Berlin, Heidelberg.

[28] Ghimire, D. and Lee, J., 2013. Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines. *Sensors*, *13*(6), pp.7714-7734.

[29] Kumar, K.M. and Srinivasan, B., 2019. Point-of-Interest Based Classification of Similar Users by Using Support Vector Machine and Status Homophily. *International Journal of Machine Learning and Computing*, *9*(5).

[30] Ghrabat, M., Ma, G., Avila, P.L.P., Jassim, M. and Jassim, S., 2019. Content-Based Image Retrieval of Color, Shape, and Texture by Using Novel multi-SVM Classifier. *International Journal of Machine Learning and Computing*, *9*(4).

[31] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3), pp.273-297.

[32] Faruqe, M.O. and Hasan, M.A.M., 2009, August. Face recognition using PCA and SVM. In *2009 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication* (pp. 97-101). IEEE.

[33] Liu, G., Mao, S. and Kim, J.H., 2019. A Mature-Tomato Detection Algorithm Using Machine Learning and Color Analysis. *Sensors*, *19*(9), p.2023.

[34] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.