

Dynamic biometric identification from multiple views using the GLBP-TOP method

Yu Wang^{a,b,c}, Xuanjing Shen^{a,b}, Haipeng Chen^{a,b,*} and Yujie Zhai^{a,b}

^aCollege of Computer Science and Technology, Jilin University, Changchun 130012, China

^bKey Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China

^cApplied Technology College Jilin University, Changchun 130012, China

Abstract. To realize effective and rapid dynamic biometric identification with low computational complexity, a video-based facial texture program that extracts local binary patterns from three orthogonal planes in the frequency domain of the Gabor transform (GLBP-TOP) was proposed. Firstly, each normalized face was transformed by Gabor wavelet to get the enhanced Gabor magnitude map, and then the LBP-TOP operator was applied to the maps to extract video texture. Finally, weighted Chi square statistics based on the Fisher Criterion were used to realize the identification. The proposed algorithm was proved effective through the biometric experiments using the Honda/UCSD database, and was robust against changes of illumination and expressions.

Keywords: Biometric identification, gabor filters, LBP-TOP, fisher criterion

1. Introduction

Biometric identification system receives extensive attention in both academic and industrial realms. A practical biometric system should have the specified recognition accuracy, speed, and specific environment. A number of biometric characteristics are in use in various applications, such as DNA, face [1], fingerprint, voice, and iris. In the last decade, the use of biological signals like EEG and Brain Waves [2] has caught much attention. And now, multimodal biometric authentication systems [3] arise rapidly. Face recognition appears to offer several advantages over other biometric methods. The applications of facial recognition range from a static, controlled “mug-shot” verification to a dynamic recognition [1]. In order to create a successful facial recognition program, the construction of a facial model is given first priority, followed by the designation of a classifier.

Dynamic video-based facial recognition [4] has many advantages compared to identification based on static images, and three-dimensional facial models obtained from video sequences and high-resolution images can improve recognition success. In a word, time and motion play a very important role in video-based facial recognition [5].

*Corresponding author: HaiPeng Chen, College of Computer Science and Technology, Jilin University, Changchun 130012, China. Tel.: 13756551186; Fax: 0431-85094780; E-mail: chenhp@jlu.edu.cn.

There are two main ideas in current video-based facial recognition: Extract the most representative frame out of the video sequences, and then use traditional methods, including the geometrical feature approach [6], model method, and statistical method, among others. The other current procedure is to regard the video as a collection of video frames, and using a 3D probabilistic model approach with the spatiotemporal data [7], design a video texture description from the imagery [8] to improve the recognition rate. The focus of video-based face recognition is to make full utilization of the spatiotemporal information of the face to overcome such difficulties as low resolution, the large-scale range of changes in illumination, and posture changes, all with low computational complexity.

This article proposes a method for facial texture description and corresponding recognition based on local Gabor binary pattern from three orthogonal planes. This method aims to effectively express video-based facial texture and perform recognition with low computational complexity, based on video frame sets. By combining Gabor transform and LBP-TOP operator, this algorithm is able to extract video-based facial texture integrated features from multiple angles, thus exhibiting improved capacity of expressing feature histograms. In addition, the proposed algorithm demonstrates robustness against changes of illumination and expression, and strong discrimination ability.

2. Framework of the proposed algorithm

A novel video-based facial representation and recognition method is proposed in this paper. Figure 1 illustrates the framework of the proposed algorithm, and the strategy is as follows:

1. Normalize the face frames extracted from the video sequence by the coordinate of the eyes.
2. Execute the Gabor method of multi-scale and multi-orientation in every face frame of the video sequence to get the amplitude image. Integrate multiple amplitude images to extract the frequency features of face frames.
3. Divide the video sequence into 8×8 blocks, extract LBP-TOP video texture from each block of video, and obtain the histogram of the statistical information for each video block, which can be treated as the spatiotemporal information.
4. Apply the cascade operation on the LBP-TOP features histogram of each video block to establish a global histogram of the entire video sequence, and then the representation from a multi-view of features can be obtained
5. Acquire the classification results by computing the similarity between the feature histogram of the query face sequence and the training face sequences. Weighted Chi square statistics based on the Fisher Criterion can be used to improve the recognition rate further.

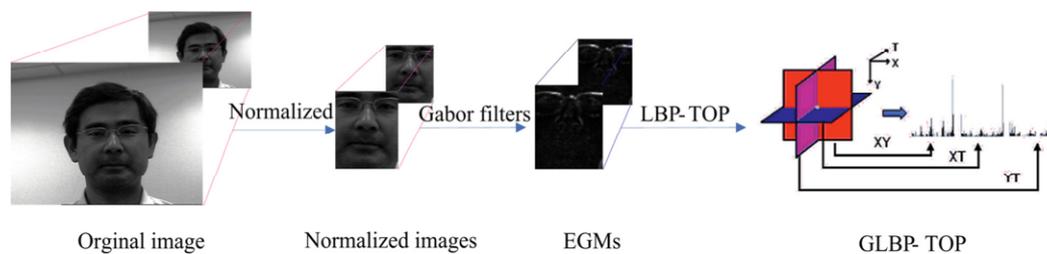


Fig. 1. The framework of the proposed algorithm.

According to Figure 1, the proposed algorithm takes into account important feature information in the frequency, spatial, and time domains simultaneously. By using the GLBP-TOP histogram, not only can the video texture information be obtained, but also the statistical information of the local area. The histogram vector of each local area in a particular order is joined together to get overall texture information of the entire facial video sequence. This video-based facial recognition method proved to be quick and effective even with illumination and expression changes, and was realized with low computational complexity.

3. Video-based facial texture representation

3.1. Enhanced Gabor magnitude images

The Gabor wavelet [9] has good success in extracting target information from local spatial and frequency domains [10].

The definition of the 2D Gabor wavelet is:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} \exp\left(-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}\right) \cdot [\exp(ik_{\mu,\nu}z) - \exp(-\frac{\sigma^2}{2})] \quad (1)$$

where μ and ν represent the orientation and scale of the Gabor kernel, $z = (x, y)$ represent a pixel of the image.

The Gabor feature of facial images can be obtained from the convolution of an image and the Gabor filter as:

$$G_{\psi, f}(x, y) = f(x, y) * \psi(z) \quad (2)$$

where $f(x, y)$ is the image to be transformed. The convolution result of the Gabor wavelet and the image is a complex response composed of real and imaginary components (filter coefficients). The amplitude is relatively stable, not rotated along with position, and reflects the energy spectrum of the image, so it is often represented as a facial feature. The amplitude is represented as:

$$M(x, y) = \sqrt{(\text{Re}(G(x, y)))^2 + (\text{Im}(G(x, y)))^2} \quad (3)$$

where $\text{Re}(G(x, y))$ represents the real part, $\text{Im}(G(x, y))$ represents the imaginary part of the convolution result. Take $m=\{0,1\}$ and $n=\{0,1,2,3\}$ amplitude texture in two scale and four orientation, and try to enhance the feature of the Gabor magnitude texture as:

$$T(x, y) = \left(\sum_{m=0}^1 \sum_{n=0}^3 M_{mn}^2(x, y) \right)^{1/2} \quad (4)$$

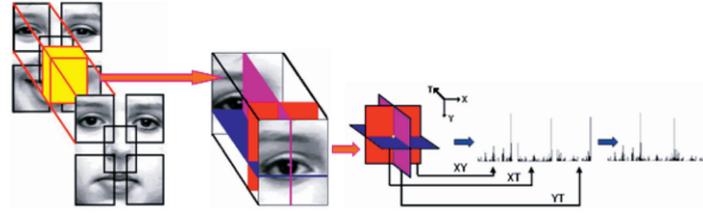


Fig. 2. Spatiotemporal texture represented by cascading LBP-TOP histograms.

3.2. Local binary patterns from three orthogonal planes

VLBP (Volume Local Binary Pattern) adds motion information to the original LBP [11-13]. It extracts the texture and motion features, and can be used for facial recognition in an image sequence. VLBP can be calculated as:

$$VLBP_{L,P,R}(x,y,t) = \sum_{p=0}^{3P+1} 2^p s(I_p - I_c) \quad (5)$$

where I_c corresponds to the gray value of the center pixel, I_p refers to gray values of $3P+2$ neighbors equally spaced pixels on a circle of radius R . These neighbors include not only the P neighbor pixels of the current frame, but also $P+1$ neighbor pixels of two frames whose time interval is L .

Figure 2 illustrates the process of the LBP-TOP operator [14,15]. A video sequence can be considered as a stack of XY planes whose center axis is the time axis T, but it ignores the fact that the image sequence can also be stacked by the XT plane with the center axis of Y axis, and stacked by the YT plane with the center of X axis. The XT and YT planes provide valuable information about the spatial and temporal transition. And the LBP-TOP operator extracts the cascading LBP features in the three orthogonal planes of XY, XT and YT, considering the change information of texture in these three directions. The motion and appearance information of the video sequences can be represented, which include spatial information (XY-LBP) and time-spatial domain information (XT-LBP and YT-LBP). The LBP-TOP operator can reduce the dimension of the histogram from 2^{3P+2} to 3×2^P with respect to the VLBP operator when the P neighbor pixels are considered in one frame, greatly reducing the dimension of the histogram, and the computational complexity.

4. Video-based face recognition

4.1. Weighted Chi square statistics

Since GLBP-TOP is a sequence composed of multiple histograms, Chi square statistics can be used to measure the similarity between two sequences of histograms. Chi square statistics are calculated as:

$$\chi^2(S, M) = \sum_{i=0}^{L-1} \frac{(S_i - M_i)^2}{S_i + M_i} \quad (6)$$

where S and M represent two histograms, and L is the dimension of the histogram.

Different weights should be given to different regions of the human face, reflecting the variations of each region, and having a positive impact on the recognition. According to the analysis above, the similarity calculation can be rewritten as:

$$\chi_w^2(S, M) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \tag{7}$$

where, w_j is the weight of j -th block in the global feature.

4.2. Weights learned based on Fisher criterion

The main idea of the Fisher criterion is to learn the weights of each region based on keeping the within-class scatter as small as possible and between-class scatter as large as possible [16,17]. For the C class problem, the similarities of the different samples from the same class form the within-class scatter, while the difference samples from different classes form the between-class scatter. Then the similarity mean and variance of the within-class can be calculated by Eqs. (8) and (9).

The within-class similarity mean is calculated as:

$$m_{I,b} = \frac{1}{C} \sum_{i=1}^C \frac{2}{N_i(N_i - 1)} \sum_{k=2}^{N_i} \sum_{j=1}^{k-1} \chi^2(S_b^{(i,j)}, M_b^{(i,k)}) \tag{8}$$

The within-class similarity variance is:

$$S_{I,b}^2 = \sum_{i=1}^C \sum_{k=2}^{N_i} \sum_{j=1}^{k-1} (\chi^2(S_b^{(i,j)}, M_b^{(i,k)}) - m_{I,b})^2 \tag{9}$$

where N_i represents the number of samples belonging to class i , and b is used to indicate the current block. Similarly, the mean and variance of the between-class similarity mean and variance can be calculated with Eqs. (10) and (11).

The between-class similarity mean can be found with:

$$m_{E,b} = \frac{2}{C(C-1)} \sum_{i=1}^{C-1} \sum_{j=i+1}^C \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} \chi^2(S_b^{(i,k)}, M_b^{(j,l)}) \tag{10}$$

The between-class similarity variance is calculated as:

$$S_{E,b}^2 = \sum_{i=1}^{C-1} \sum_{j=i+1}^C \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} (\chi^2(S_b^{(i,k)}, M_b^{(j,l)}) - m_{E,b})^2 \quad (11)$$

And finally, the weight of block b can be calculated as:

$$w_b = \frac{(m_{I,b} - m_{E,b})^2}{S_{I,b}^2 + S_{E,b}^2} \quad (12)$$

Figure 3 illustrates the results of the weights learned based on the Fisher criterion. Darker intensity means smaller weight and brighter intensity means larger weight. If the difference between the within-class similarity mean and the between-class is large, and both the variances are small, then the block has a better ability to identify, and a corresponding relatively large weight should be assigned to this block, otherwise the weight should be small.

5. The experiment and analysis

The Honda/UCSD database [18] has been widely used in video-based facial recognition. In the Honda/UCSD database, the training set contains 20 videos of 20 people, and the test set contains 39 videos of 19 people. In addition, since the videos in the Honda/UCSD database were collected at different time periods, and the expression and illumination vary, the robustness of the proposed method against changes of illumination and expressions could be proved. The front facial images extracted from the video sequences were normalized to 183 pixels \times 229 pixels. For this database, accuracy is the most intuitive criterion for the evaluation of the algorithm's performance.

When the length of the video sequence was 7 frames, the results of VLBP [14], LBP-TOP [14], Extend VLBP (EVLBP) [19], Gabor filter-based VLBP (GVLBP), Gabor filter-based LBP-TOP (GLBP-TOP) and Weighted Gabor filter-based LBP-TOP (WGLBP-TOP) are presented in Table 1 and Figure 4.

In LBP-TOP_{P_{XY}, P_{XT}, P_{YT}, R_X, R_T, R_Y} operator description, P_{XY}, P_{XT}, and P_{YT} represented the number of neighboring pixels of the XY plane, the XT plane, and the YT plane, respectively, and R_X, R_T, and R_Y

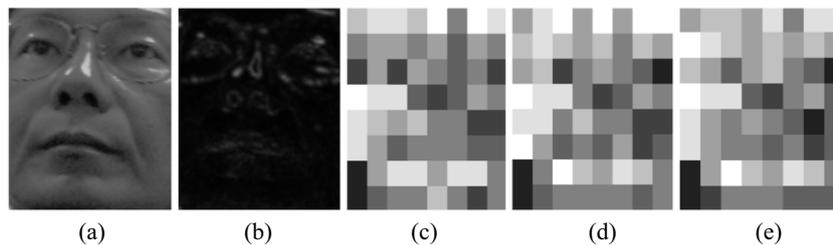


Fig. 3. Weights learned based on Fisher Criterion should be normalized to [1-8]. (a) Facial image (b) Gabor face (c) Weights for 5 frames (d) Weights for 7 frames (e) Weights for 9 frames.

represented the circle radius of neighbors in different planes. The classification was driven by the nearest neighbor classifier (1NN) based on Chi square statistics. And in WGLBP-TOP, the weighted Chi square statistics based on the Fisher Criterion were used in the matching phase.

Compared to the simple methods adopting just the VLBP or LBP-TOP facial representations, the proposed method is an effective video-based facial recognition. The faces in the tested videos were collected at different time periods. The comparison of the proposed algorithm with other algorithms is shown in Figure 4, and it validates the effectiveness of the proposed method. In addition, the proposed method also demonstrated good robustness against changes of illumination and expressions.

Table 1
The results of different algorithms

Algorithm	Size of Frames Set	Accuracy (%)
VLBP _{1,4,1} +1NN[14]	7	38.5
EVLBP _{1,(4,1,1),1} +1NN[19]		43.6
GVLBP _{1,4,1} +1NN		53.8
LBP-TOP _{8,8,8,1,1,1} +1NN[14]		56.4
GLBP-TOP _{8,8,8,1,1,1} +1NN		66.7
WGLBP-TOP _{8,8,8,1,1,1} +1NN		69.2

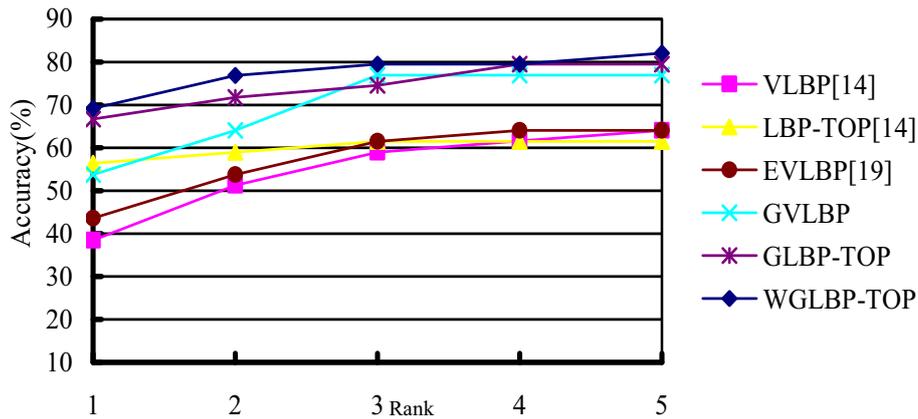


Fig. 4. CMC curves of different algorithms on Honda/UCSD.

Table 2
The results of different algorithms with different sizes of frame sets

Algorithm	Size of Frames set	Accuracy (%)	Times(ms)
GLBP-TOP _{8,8,8,1,1,1} +1NN	5	66.7	258 ± 8
	7	66.7	414 ± 8
	9	69.2	578 ± 16
WGLBP-TOP _{8,8,8,1,1,1} +1NN	5	66.7	265 ± 15
	7	69.2	422 ± 25
	9	74.4	578 ± 10

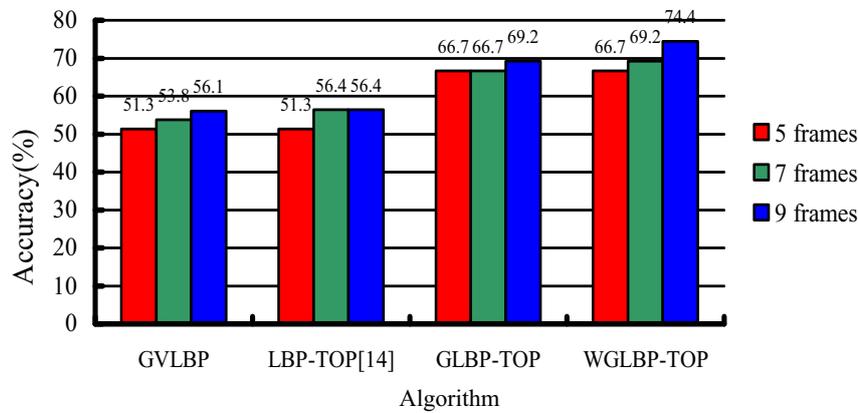


Fig. 5. The results of different algorithms with different sizes of frame sets.

In the proposed algorithm, the size of the video frame sets had great influence on the recognition rate, as shown in Table 2 and Figure 5. If more video frames were considered, a higher recognition rate was achieved, to a certain extent, although the time consumption increased.

The results above reveal that the longer the video sequence is, the better the results will be. This can be explained by the fact that the longer the video sequence is, the more probability of the system learning the facial dynamics of the individual. While the available information increased, the noise that influenced the recognition performance increased as well, and increased the time spent to acquire a match. Continuously increasing the video frame sets did not necessarily improve the recognition rate. However, the proposed method ensured a better recognition performance with low computational complexity.

6. Conclusion

A robust dynamic biometric identification algorithm from multiple views (moving images) using GLBP-TOP method was presented. It stood up well to light variations and expression changes, and other inconsistencies and interferences. The weighted learning algorithm based on the Fisher Criterion was also used in the matching and recognition phase to further improve the system's recognition performance. There is still room for improvement, including designating a more efficient classifier, but the proposed system showed advantages over current identifiers.

Acknowledgment

This work was supported by the National Science Foundation for Youths with Grant No. 61305046, supported by Jilin Province Natural Science Foundation with Grant No. 20140101193JC, and supported by Jilin Province Science Foundation for Youths with Grant No. 20130522117JH.

References

- [1] A.K. Jain and A.R.S Prabhakar, An introduction to biometric recognition, *IEEE Transactions on Circuits and Systems for Video Technology* **14** (2004), 4–20.
- [2] P. Campisi and D.L. Rocca, Brain waves for automatic biometric-based user recognition, *IEEE Transactions on Information Forensics and Security* **9** (2014), 782–800.
- [3] R. Snelick, U. Uludag, A. Mink, M. Indovina and A. Jain, Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** (2005), 450–455.
- [4] J.R. Barr, K.W. Bowyer, P.J. Flynn and S. Biswas, Face recognition from video: A review, *International Journal of Pattern Recognition and Artificial Intelligence* **26** (2012), 1266002-1–1266002-53.
- [5] F. Ahmad and A. Najam, Video-based face classification approach a survey, *International Conference on Robotics and Artificial Intelligence*, 2012, 179–186.
- [6] P. Belhumer, P. Hespanha and D. Kriegman, Eigenfaces vs fisherfaces: recognition using class specific linear projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19** (1997), 711–720.
- [7] D. Huang, C.F. Shan and M. Ardabilian et al., 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** (2011), 765–781.
- [8] G.Y. Zhao and T. Ahonen et al., Rotation-invariant image and video description with local binary pattern features, *IEEE Transactions on Image Processing* **21** (2012), 1465–1477.
- [9] M. Porat and Y. Zeevi, The generalized Gabor scheme of image representation in biological and machine vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10** (1988), 452–468.
- [10] C.J. Liu and H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, *IEEE Transactions on Image Processing* **11** (2002), 467–476.
- [11] T. Ojala, M. Pietikäinen and T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** (2002), 971–987.
- [12] A. Timo, H. Abdenour and P. Matti, Face recognition with local binary patterns, *European Conference on Computer Vision* **3021** (2004), 469–481.
- [13] T. Ahonen, A. Hadid and M. Pietikäinen, Face description with local binary patterns: application to face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28** (2006), 2037–2041.
- [14] G. Zhao and M. Pietikäinen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29** (2007), 915–928.
- [15] C.H. Chan, M.A. Tahir, J. Kittler and M. Pietikäinen, Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35** (2013), 1164–1177.
- [16] M. Taini, G.Y. Zhao and M. Pietikäinen, Weight-based facial expression recognition from near-infrared video sequences, *16th Scandinavian Conference on Image Analysis* **5575** (2009), 239–248.

- [17] X.H. Huang, G.Y. Zhao, M. Pietikäinen and W.M. Zheng, Expression recognition in videos using a weighted component-based feature descriptor, 17th Scandinavian Conference on Image Analysis **6688** (2011), 569–578.
- [18] K.C. Lee, J. Ho, M.H. Yang and D. Kriegman, Video-based face recognition using probabilistic appearance manifolds, IEEE International Conference on Computer Vision and Pattern Recognition, 2003, 313–320.
- [19] A. Hadid and M. Pietikäinen, Combining appearance and motion for face and gender recognition from videos, Pattern Recognition **42** (2009), 2818–2827.