Feature Extraction Method for Lip-reading under Variant Lighting Conditions

Xinjun Ma, Hongjun Zhang and Yuanyuan Li
Harbin Institute of Technology
Shenzhen, China
aumaxj@126.com, a901215@126.com, 2209157781@qq.com

ABSTRACT

Automatic lip-reading is a technique of understanding the uttered speech by visually interpreting the lip movement of the speaker. with the development of the lip-reading, more and more related technologies are proposed. However, the current research of lip-reading is mainly conducted under the ideal lighting conditions and there is few researchers focus on the lip-reading technique under variant lighting conditions. For this problem, this paper proposes a new method of lip feature extraction under variant lighting conditions. The method consists of a preprocessing chain of illumination normalization and improved LBP features, which can improve the recognition rate of lip-reading under variant lighting conditions from two aspects. Experiments show that the lip feature extraction method proposed by this paper is effective.

CCS Concepts

• CCS → Computing methodologies → Machine learning → Machine learning algorithms → Feature selection.

Keywords

Lip-reading; Illumination Normalization; Improved LBP.

1. INTRODUCTION

Automatic lip-reading is a technique of understanding the uttered speech by visually interpreting the lip movement of the speaker. By observing the movements of the speaker's lips and even the changes of the teeth and tongue, we can get a lot of information that the speaker wants to express [1-2]. Lip-reading technology can be applied widely in many fields, such as human-computer interaction, identity recognition, disabled auxiliary etc [6-7]. In recent years, in the field of artificial intelligence and pattern recognition there are more and more papers research on lip-reading recognition algorithm, and many methods have been proposed successfully [3-5].

At present in lip reading study, lip feature extraction is a very important problem. Many existing papers have proposed different algorithms. The traditional visual feature extraction method can be divided into 3 types: geometric feature extraction method, appearance feature extraction method and mixed feature extraction method. Geometric feature extraction method extracts geometric information of such as the height, width and area of the lip. Paper [8] uses the convex hull method to fit the lip contour, and on this basis to extract the height, width, area and perimeter of lips, finally using TP-MDTW classifier to recognize 10 English digital. The highest recognition rate reached 72%. Appearance feature extraction method makes the gray pixel of lip region as a high dimensional vector. For this method, lip feature is got directly from the ROI after lip localization. The commonly used methods of this kind are mainly PCA, DCT, LDA, DWT, etc. Recently, a multi-position lip-reading recognition system developed by Lucey et al used visual feature extraction methods based on DCT, and using linear discriminant analysis (LDA) to reduce the feature dimension and finally carry out experiment in a variety of head pose angles, the highest recognition rate reached 84.61%[9]. This feature extraction method is simple and fast, but requires the division of needle lip area in the image sequences to maintain strict consistency, which is difficult to guarantee in the case of non-manual operation.

The mixed feature method is combined with different types of lip features to describe the change of lip movement more comprehensively and accurately. One of the methods is active appearance mode (AAM). Paper [10] adopts the model method to fit the outer contour of lip based on the 46 lip coordinates from AAM and fit inner lip contour by self-defined pixel module matching so that to get feature model which can describe the lip movement comprehensively and finally by conduct experiment for 30 Korean isolated vocabulary, the highest recognition rate reached 92.67%, which is higher the traditional AAM method. Mixed feature methods extract lip feature from different aspects, and get a more complete description of lip movements. However, the complexity of the method is improved at the same time.

For the above feature extraction methods, the most widely used is the appearance feature extraction method. There are also a lot of papers which show that this method has the highest recognition rate of lip feature extraction method in several ways. So this paper will focus on the research of this method. Research on these algorithms mainly focus on the positive ideal lighting conditions, but the actual lip-reading recognition system will work in the complex application environment. The change of external lighting conditions and face posture, bearded and partly obscured lip and other issues will have bad influence on lip feature extraction and recognition in different degrees. Among them, illumination change in the practical application has the most badly impact, many algorithms which have good effect in the ideal lighting conditions perform badly in the variant lighting conditions.

At present, there are a lot of methods and papers to deal with the problem of face recognition under variant lighting conditions. However, papers focusing on influence to lip-reading caused by

ICMLC 2017, February 24-26, 2017, Singapore, Singapore.
© 2017 ACM. ISBN 978-1-4503-4817-1/17/02…$15.00
DOI: http://dx.doi.org/10.1145/3055635.3056576

variant lighting conditions are very few and even close to the blank. Therefore, this paper is devoted to the study of lip-reading recognition under variant lighting conditions. Considering the influence caused by the external lighting conditions, this paper improved the method of lip feature extraction from two aspects and proposed a new method of lip feature extraction based on preprocessing chain of illumination normalization and improved LBP features.

Section 2 of the paper describes the specific implementation of the preprocessing chain of illumination normalization in detail and the corresponding experimental results. In Section 3, the algorithm of lip feature extraction based on improved LBP illumination invariant feature is discussed. Section 4 gives the structure of lip-reading system designed in this paper. Besides, the corresponding method of lip localization and recognition are also introduced. Finally, the experiment based on MATLAB is carried out, which proves the rationality and effectiveness of the proposed algorithm.

2. PREPROCESSING CHAIN OF ILLUMINATION NORMALIZATION

Aiming at eliminating the interference caused by variant lighting conditions, this paper proposed a preprocessing chain of illumination normalization. The preprocessing algorithm is applied to remove the influence of external illumination noise before the lip feature extraction. The whole preprocessing algorithm consists of four steps, which are median filtering, gamma correction, multi-scale retinex filtering and contrast equalization. Each step will be a certain degree of compensation for the light interference.

Median filtering, in the process of formation and transmission of the video images captured by the camera, it often leads to a large number of pulse noises because of external noise interference. As for analog signals, the effect of impulse noise is very little, but in the transmission of digital signals, impulse noise will greatly affect the quality of the image. Median filter can remove the impulse noise from the image, and it can also protect some of the details in the image. Therefore, a median filtering template of 3*3 is used to filter the input image to remove impulse noise.

Gamma correction influenced by external illumination, there is always uneven distribution of light in the ROI of lip. The most typical case is that because of light reflection, part of the area is too bright and the other part is too dark. In order to make gray correction to the lip image aiming at above situation in the image preprocessing stage, this paper further carries on the gamma correction to improve the lip image illumination distribution after making median filtering to the lip area. The formula for Gamma correction is shown in equation 1. The gamma value selected in this paper is 1/2.2.

$$I'_{(x,y)} = \left(\frac{I_{(x,y)}}{255} \right)^{\gamma} \times 255 \, , \; \gamma \in (0,1) \tag{1}$$

Multi-scale Retinex Filter Multi-scale Retinex filter(MSR) is composed of the most basic Single Scale Retinex filter. It is considered that images are composed of the incident and reflected components, as shown in equation 2. Therefore, the use of MSR for image filtering is essentially to calculate the incident component of an image accurately and eliminate the component in the original image. Because it is a singular problem to calculate the incident component in the input image directly, it can only be approximated by the mathematical method. This paper uses Gauss surround function to accomplish this task, the specific methods are as follows:

$$S(x,y) = R(x,y)L(x,y) \tag{2}$$

$$\begin{aligned} \log R(x,y) &= \log[S(x,y)/L(x,y)] \\ &= \log S(x,y) - \log[S(x,y) \otimes G(x,y)] \end{aligned} \tag{3}$$

Among them, $G(x,y) = \lambda \exp(-\frac{x^2+y^2}{\sigma^2})$, σ is scale function of Gauss's surround function. $G(x,y)$ satisfy that $\iint G(x,y)dxdy = 1$, λ is a constant to normalize $\iint G(x,y)dxdy = 1$. The benefits of using a logarithmic operation is that it can convert the multiplication and division relationship to addition and subtraction relationship in order to simplify the operation, on the other hand, logarithm itself also has certain light filtering function. The structure of the filtering algorithm is shown in Figure 1.

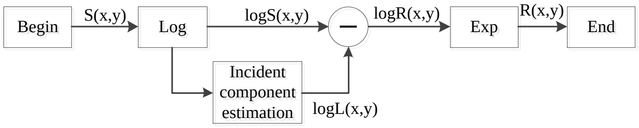


Figure 1. Flow chart of Rextinex Filtering algorithm..

σ is the most important parameter in the use of the Gauss surround function in this paper. When σ is small, ability of dynamic range of gray compression is strong, which can better highlight the lip image details, but also causes the image distortion to some extent; σ is large, the lip image has high fidelity, but also correspondingly reduced for the ability of dynamic range of gray compression. In order to make up for this defect, this paper uses the MSR algorithm to filter the image, and the formula of the logarithm field is shown in the form of 4.

$$\log R(x,y) = \sum_{k=1}^{K} \omega_k \{\log S(x,y) - \log[S(x,y) \otimes G_k(x,y)]\} \tag{4}$$

Here, in order to ensure that the filter function has the advantages of both the single scale Retinex filter high, medium and low three scales, the K value is 3, and the filter with three scales has the same weight, that is, $\omega_1 = \omega_2 = \omega_3 = 1/3$. After repeated experiments, when the 3 scale factors σ are respectively as15, 80 and 250, the filter achieves the best filtering effect.

$$I(x,y) = \frac{I(x,y)}{(mean(|I(x',y')|^a))^{1/a}} \tag{5}$$

$$I(x,y) = \frac{I(x,y)}{(mean(\min(\tau, |I(x',y')|)^a)^{1/a}} \tag{6}$$

$$I(x,y) = \tau \frac{2}{\pi} arc\tan(I(x,y)/\tau) \tag{7}$$

Contrast equalization After MSR filtering, lip image illumination situation has been improved significantly. After the filtering, image gray distributes in different grayscale range, and the grayscale range is very narrow, thus the final step in the proposed light preprocessing algorithm can improve lip image light distribution by contrast equalization of lip image. The contrast equalization operation for it is shown in equations of 5-7.Here, a is the image compression factor, which can adjust the dynamic range of gray lip image effectively. τ is used to restrict the large gray value threshold. In this paper, we take a=0.2, τ =8. The formula 7 is an arctangent transformation, which is a nonlinear transformation. The image can be normalized to the range $(-\tau,\tau)$ effectively by applying arctangent transform to lip image.

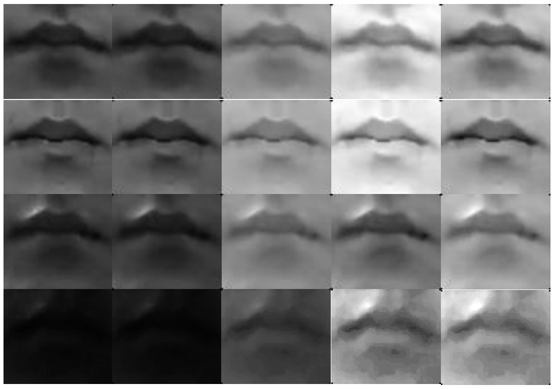


Figure 2. The effect of the illumination normalization algorithm.

Figure 2 shows the effect drawing after applying the proposed light preprocessing algorithm .In each row, it is a different kind of light situations. The first image of each line is a lip image without any processing, and the remaining 4 images are effects after being processed by different steps. It can be seen from Figure 2 that the proposed light preprocessing algorithm is effective.

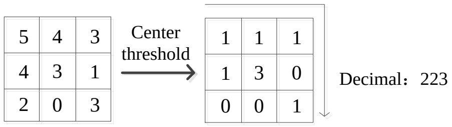
3. ILLUMINATION INVARIANT FEATURE EXTRACTION BASED ON IMPROVED LBP

The method based on pixel can be divided into two categories: global feature extraction method and local feature extraction method. The global feature describes the overall information of the object, and it can also be used to represent the object roughly. However when it comes to the illumination change, robustness is poor. what's more, it cannot eliminate the effects caused by local changes and is vulnerable to external illumination variation. On the contrary, local features reflect and capture local details of the object, and it can use the local information in the region to represent a single pixel feature, so when it comes to the illumination change, robustness is good and it is not easily affected by external interference illumination.

At present, the most representative of the local feature extraction method is Local Binary Patterns. Because feature extraction method has high resolution and is insensitive to gray scale change, it has been widely applied in face recognition, and it is proved to have good illumination robustness under variable illumination conditions. However, the application of the method is rarely mentioned in the field of lip-reading. Therefore, this excellent feature extraction method can be applied to the research of the lip-reading, and we can use it to improve the effect of lip-reading recognition under variable illumination.

3.1 Local Binary Patterns

LBP operator was first proposed by Ojala et al. The basic operation unit of the method is composed of a 9 pixel module, and takes the center pixel as threshold, and processes 8 adjacent pixels around it. If the pixel value is more than the threshold，the result is 1, and otherwise it is 0. Then the 8 values that are connected in sequence make up of LBP encoding value, finally when the binary LBP code is converted to decimal number, we can get the LBP mode value of the center pixel, the specific process is shown in figure 3. Its corresponding mathematical expressions are shown in the formula 8.



5	4	3
4	3	1
2	0	3

1	1	1
1	3	0
0	0	1

Figure 3. Schematic diagram of the basic LBP operator

Although the basic LBP operator has made some recognition results, there are some defects, which is the fixed sampling area and the limited number of sampling points. After that, Ojala and others made further extensions for the basic LBP operator. In the extended LBP model, the original square template is replaced by a circular template with a radius of R. P sampling points in circular template edge are compared with the pixel value located in the center of the circle, which can generate the LBP model value. When the sampling point is located at the center of the pixel points, the bilinear interpolation method is used. This effectively enhances the sampling ability of LBP operator and we can sample the neighborhood pixels of the image, and be able to extract more effective local features, and the recognition ability of the extracted features is significantly better than the basic LBP features. Figure 4 is LBP sampling diagrams with different radii and sampling points.

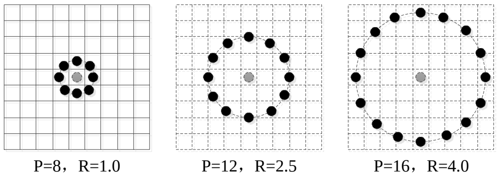


Figure 4. 3 circular LBP with different radii.

The coordinates of the center pixel is (x_c, y_c) , and the coordinates of the sample points on the edge of the circular template can be obtained by the following formula:

$$x_p = x_c + R\cos(2\pi p / P) \tag{8}$$

$$y_p = y_c + R\sin(2\pi p / P) \tag{9}$$

If the coordinate value is not an integer, the interpolation method is needed to calculate the corresponding pixel value. In this paper，we use the bilinear interpolation method. The mathematical expression of the circular LBP operator is shown in the formula 10.

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{10}$$

$$s(x,y) = \{\begin{matrix} 1 & x \geq 0 \\ 0 & x < 0 \end{matrix} \tag{11}$$

Among them, g_p is the first p pixel point on the edge of the circular template, and g_c is the center pixel of the circular template. According to the formula 10, we can get the 3 kinds of LBP mode in Figure 4, it can be recorded respectively as $LBP_{8,1}$, $LBP_{12,2.5}$ and $LBP_{16,4}$.

After the extraction of lip LBP characteristics, and it will not be directly as a feature, because this feature vector has high dimension, and will directly slow down the lip-reading recognition speed. In order to solve this problem, we adopt the unified model of LBP in this paper. LBP unified model refers that the change of 0 and 1 in LBP binary code is no more than 2 times, for that LBP code is a circular distribution. For LBP unified model operator, it is denoted by $LBP_{P,R}^{u2}$. The unified mode of LBP contains most of the information of the lip image, and avoids a lot of noise information in the unified model of lip image, so lip-reading recognition rate is higher if we use LBP unified model.

In order to determine the optimal LBP operator for this topic of lip reading, in this paper, different LBP parameters are used to do a number of group identification tests, and the test results are shown in Table 1. Among them, the recognition rate of each LBP parameter is the average of the 20 test results. At the same time, in order to evaluate the stability of the recognition results, the standard deviation of the recognition rate obtained by different LBP parameters is also calculated, which is based on the recognition rate. It is not difficult to see from table 1, when P=8, R=3, the recognition rate is the highest, and the standard deviation is minimum, and the stability of the identification is also the best, so we will use this set of parameters to carry on the follow-up experiment in this paper.

Table 1. Experimental results corresponding to different LBP parameters

LBP operator	Recognition rate(%)	Standard deviation
$LBP_{4,1}^{u2}$	70.05	0.025
$LBP_{4,1}^{u2}$	71.32	0.029
$LBP_{4,1}^{u2}$	63.91	0.018
$LBP_{4,1}^{u2}$	76.30	0.022
$LBP_{4,1}^{u2}$	77.57	0.021
$LBP_{4,1}^{u2}$	76.60	0.021
$LBP_{4,1}^{u2}$	77.98	0.015
$LBP_{4,1}^{u2}$	77.94	0.019
$LBP_{4,1}^{u2}$	76.16	0.025
$LBP_{4,1}^{u2}$	77.72	0.019
$LBP_{4,1}^{u2}$	77.63	0.022
$LBP_{4,1}^{u2}$	76.66	0.021
$LBP_{4,1}^{u2}$	76.60	0.021
$LBP_{4,1}^{u2}$	78.47	0.023

3.2 Improved LBP Feature Extraction Method

The above-mentioned LBP feature extraction is performed in the whole lip image, and it results in missing microstructure information of describing lip image in final LBP histogram vector, to some extent which also reduces the recognition rate of the lip-reading system. LBP histogram represents first order statistical characteristic of image gray value after LBP coding. However, when only extracting a single LBP histogram, we can't get position distribution information of each gray level of the image, so there is no way to describe the microstructure information of lip image. Usually in a lip image, the difference of local information in different regions is relatively large.

We can get LBP coding map of previous lip image by applying LBP feature extraction operator to the whole lip image. on this basis, we make histogram statistics for the LBP image, and generate LBP histogram. However, it usually losts the local difference information of the lip image in LBP histogram generated by the whole lip image. Therefore, this article adopts the method is that partition the lip image, and extract the LBP histogram of each partition, so you can increase the microstructure information of lip image when you retain its local information. Based on this idea, we can divide lip image whose size is M×N into the partitions whose size is A×B, so the size of each partition is (M/A)×(N/B). Then we can calculate LBP histogram of the partitions. Test different blocked modes in this paper， we finally chose the block method, the specific process as shown in figure 5.

The LBP feature extraction method based on partition can extract the local lip feature image well, so it has certain robustness for the external light changes. But this method cannot extract global features of lip images, and lip features can't get a complete description. In order to extract the most representative features of the lip image further, we combine PCA technology with LBP feature extraction method, thus we not only can extract the local features of the lip image and but also can extract the global features. Combination of them can improve the lip-reading recognition rate further. Another advantage of using PCA method is that it can reduce the dimension of the lip feature. In this paper, corresponding image sequences of each number collected in lip-reading database are more than 25 frames, thus corresponding lip feature of a number is at least 25 times as much as the single frame lip image. The PCA method is used to reduce the dimension of the LBP block histogram vector, which not only can improve the recognition rate of lip-reading system and filter interference brought by Illumination change, but also can improve the recognition rate of the lip-reading system.

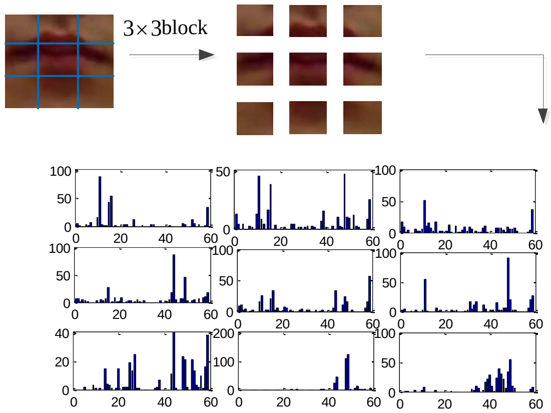


Figure 5. Sketch of LBP histogram vector feature extraction

4. SYSTEM STRUCTURE OF LIP-READING

Recognition of lip-reading system presented by this paper is mainly for isolated words. The framework of the system is shown in figure 6. From the figure we can see that the lip-reading system mainly consists of four parts, respectively for lip segmentation and to illumination preprocessing, lip feature extraction and lip-reading model training and recognition.

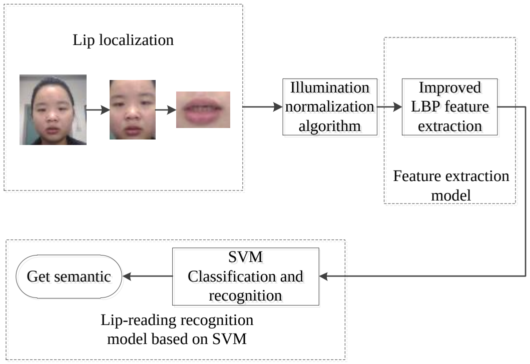

Lip localization
Illumination normalization algorithm
Improved LBP feature extraction
Feature extraction model
SVM Classification and recognition
Get semantic
Lip-reading recognition model based on SVM

Figure 6. The framework of lip reading recognition system proposed in this paper.

Among them, for the lip localization, this article adopts the method that locates face before the lip. For face location, we use Haar-like features and Adaboost algorithm which is widely used at present. The algorithm is robust to illumination. Besides, the position of the face in the scene is fixed and the proportion of the face in the screen is large, so the method can completely meet the requirements. In the face image has been obtained, we use the lip locating method put forward before by the research group to locate lip region [11]. The recognition method used in this paper is SVM. When the training sample is little, the classification effect of SVM is better than other classification methods. From so on, in lip-reading recognition, hidden markov model (HMM) is widely used, but because HMM requires a large number of samples for training, and the sample data in this paper is small, so HMM is not the best way. And some literature also shows that SVM is better than HMM in small database [12]. Therefore this paper will use SVM for classification. For that the input feature vector dimension of SVM need be fixed, here we method proposed by [13] to solve it.

5. EXPERIMENTAL RESULTS AND ANALYSIS

5.1 The Establishment of Lip-reading Database

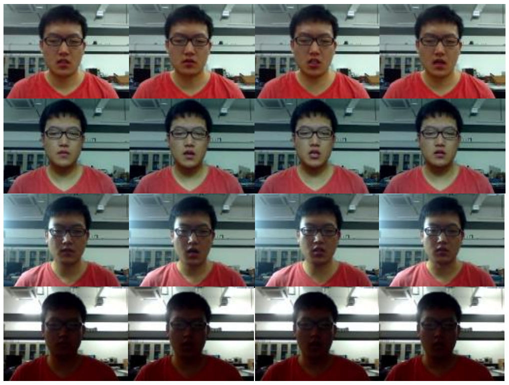


Figure 7. The sample of the lip-reading database of variant lighting conditions built in this paper.

This database is based on the corpus of Chinese pronunciation 0-9 a total of 10 sets of figures. In this paper, the volunteers in 4 different light conditions were made to repeat the number of 0-9 20 times. The subset of 1 is increased 5 times, in order to better study natural light of lip-reading. This database contains a total of 850 samples. Image capture resolution is set to 480*640. The video frame rate is 30fps. The corresponding audio sampling rate is 4.8 kHz. Figure 7 is the sample of the database built in this paper.

The four line images from top to down in figure 7 respectively are subset 1、subset 2、subset 3、subset 4. Among them, subset 1 can be regarded as a good subset of natural light, a subset 2 can be regarded as the external light excess, subset 3 can be seen as a subset of light distribution, and subset 4 can be regarded as the external light dark conditions. It can be seen from the figure that the lip-reading database built by this paper fully reflects the external variant lighting conditions of the changes, which can represent the actual environment lip-reading system has to face.

5.2 Experimental Result

Based on the lip-reading database built in this paper, by using the SVM recognition method, this paper carries out four different experiments to verify the effectiveness of the proposed illumination normalization algorithm and the improved LBP feature extraction algorithm in this paper. Among them, the first group was compared with the effect of different feature extraction methods under natural light illumination. By comparing the proposed lip feature extraction algorithm with the commonly used PCA, DCT and LBP three methods, it can reflect the recognition effect of lip-reading system proposed in this paper to digital. The test using the subset 1 to do experiments, each selected 5 samples of subset 1 randomly for training, the remaining 20 samples for identification test. Each feature extraction method was tested for 20 consecutive times, and finally was taken the average value, and thus get the recognition effect for different feature extraction method for each figure, test results as shown in Figure 8.

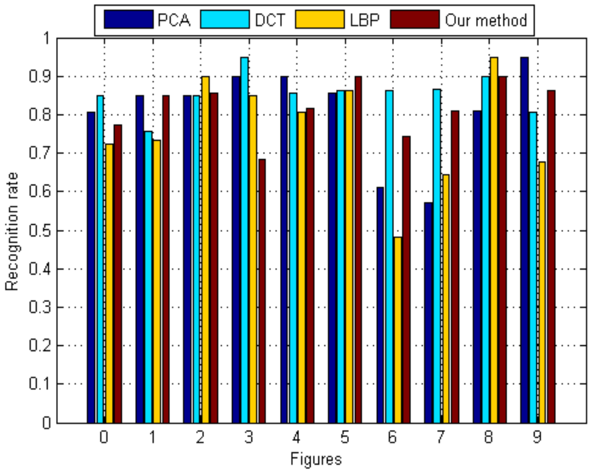


Figure 8. Comparison of recognition rate of different feature extraction methods under natural illumination.

As can be seen from Figure 8, under the natural light conditions, the DCT method is the highest recognition, better than other methods. And the method proposed in this paper is comparable to the recognition rate of PCA, and the recognition effect of LBP method is the worst. At the same time for different numbers, identification of 6 and 7 was the lowest, 5 and 8 recognition rate is the highest, which is due to the pronunciation mouth sequence of 5 and 8 is unique compared to other digital. Overall, the performance of lip reading recognition system under natural light conditions performs well, the average recognition rate reached about 81.25%.

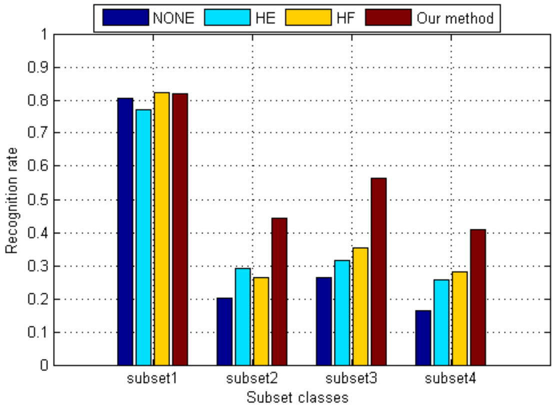


Figure 9. Recognition rate comparison of different illumination preprocessing algorithms.

In the second set of experiments, we use the improved LBP lip feature extraction method in this paper to compare the recognition results under different illumination preprocessing algorithms. The experiment was used to train the SVM model with 5 samples randomly selected from subset 1, the test samples were remaining 20 samples of the subset 1 and all subset 2, subset 3 and subset 4, respectively. The experimental results are shown in Figure 9. Among them, NONE represents that it does not use any of the illumination preprocessing algorithm, HE represents the histogram equalization, HF represents homomorphic filtering. It can be seen from Figure 9, when the light conditions change, compared to the direct extraction of lip features using illumination preprocessing algorithm will improve the recognition rate of lip-reading. Among them, the effect of histogram equalization and homomorphic filtering is equivalent, the algorithm proposed by this paper is the best. However, it is worth noting that compared to natural light conditions, when the lighting conditions change, lip-reading recognition rate down a lot. Although the compensation measures are adopted in this paper aiming at the effect caused by variant lighting conditions, but the recognition rate is still lower than that of constant illumination environment.

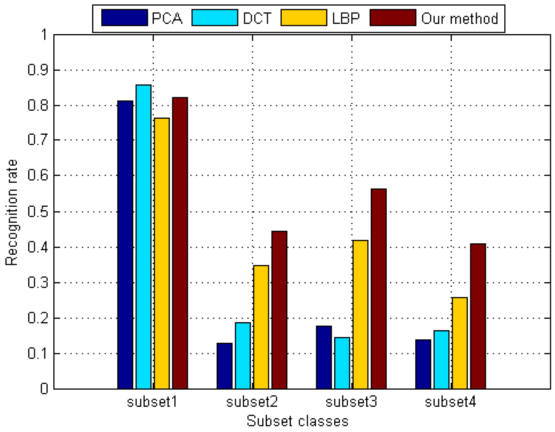


Figure 10. Comparison of recognition rate of different lip feature extraction methods.

In the third set of experiments, we adopt the illumination preprocessing algorithm proposed by this paper to compare the recognition rate of different lip feature extraction algorithm. The experimental procedure is the same with the second groups, and the experimental results are shown in Figure 10. As can be seen from Figure 10, although the recognition rate of PCA, DCT and other feature extraction methods in subset 1 is higher, but for other subsets, the recognition rate is much lower than that of LBP and the method proposed in this paper. Thus, LBP is indeed a feature extraction method of illumination robust, and the improvement for feature extraction of LBP makes the lip-reading recognition further increases to a certain extent. In contrast, the traditional feature extraction method is vulnerable to the impact of external illumination changes, the robustness is poor.

6. CONCLUSION

According to the effects caused by variant lighting conditions on lip-reading recognition result, we put forward a new method for extracting lip features. The method consists of a illumination preprocessing chain and a light invariant feature extraction operator. In the part of removing illumination influence, this paper combines the advantages of different preprocessing algorithms, designing a lighting preprocessing chain and proved its effectiveness through experiments. Based on the traditional LBP operator, this paper proposes an improved LBP method for the invariant lip feature extraction. Finally, the experiments show that the proposed lighting robust lip feature extraction algorithm is lower than the traditional pixel based feature extraction method in natural light condition for the recognition rate, but the recognition effect is better than other feature extraction methods under the variant lighting conditions.

7. ACKNOWLEDGMENTS

This work was supported in part by Shenzhen Ba-sic Science Research Foundation of China under Grant JCYJ20150513151706580.

8. REFERENCES

[1] Zhao, G., Barnard, M., & Pietikainen, M. (2009). Lipreading with local spatiotemporal descriptors. *IEEE Transactions on Multimedia, 11*(7), 1254-1265.

[2] Aleksic, P. S., & Katsaggelos, A. K. (2006). Audio-visual biometrics.*Proceedings of the IEEE, 94*(11), 2025-2044.

[3] Almajai, I., Cox, S., Harvey, R., & Lan, Y. (2016). Improved Speaker Independent Lipreading using Speaker Adaptive Training and Deep Neural Networks. *ICASSP*.

[4] Bear, H. L., Harvey, R. W., & Lan, Y. (2015). Finding phonemes: improving machine lip-reading. *Joint Conference on Facial Analysis, Animation and Auditory-Visual Speech Processing*.

[5] ALan, Y., Harvey, R., Theobald, B. J., Bowden, R., & Ong, E. J. (2010). Improving visual features for lip-reading. *International Conference on Auditory-Visual Speech Processing* (pp.142-147).

[6] Bakry, A., & Elgammal, A. (2013). Mkpls: manifold kernel partial least squares for lipreading and speaker identification. 684-691.

[7] Zhao, G., & Pietikäinen, M. (2013). *Visual Speaker Identification with Spatiotemporal Directional Features*. *Image Analysis and Recognition*. Springer Berlin Heidelberg.

[8] M.Z. Ibrahim, & D.J. Mulvaney. (2015). Geometrical-based lip-reading using template probabilistic multi-dimension dynamic time warping ☆.*Journal of Visual Communication & Image Representation, 30*(C), 219-233.

[9] Lucey, P., & Sridharan, S. (2008). A visual front-end for a continuous pose-invariant lipreading system. *International Conference on Signal Processing and Communication Systems* (pp.1-6). IEEE.

[10] Shin, J., Jin, L., & Kim, D. (2011). Real-time lip reading system for isolated korean word recognition. *Pattern Recognition, 44*(3), 559-571.

[11] Ma X, Zhang H. Lip segmentation algorithm based on bi-color space[C]// *Control Conference. IEEE*, 2015.

[12] Morade, S. S., & Patnaik, S. (2015). Comparison of classifiers for lip reading with cuave and tulips database. *Optik - International Journal for Light and Electron Optics, 126*(24), 5753-5761.

[13] Shi, Xiaoxing. "A New Time Wrapping Algorithm and Its Application on Neural Network Based Speech Recognition." Journal of Southeast Univwrsity (1999).