

Facial Expression Recognition Based on Local Features and Monogenic Binary Coding

Zhangbao Chen

School of Electronic and Electrical Engineering, Bengbu University, Bengbu, Anhui 233030, China

Corresponding address: No. 1866, Caoshan Road, Bengbu, Anhui 233030, China

E-mail: zbczbao@126.com

Keywords: monogenic binary coding, facial expression recognition, local features, Japanese female facial expression

Received: March 11, 2019

Fast developing facial expression recognition is one of the significant recognition technologies for biological features with high applied value. In this study, a monogenic binary coding algorithm was considered to illustrate a good matching with local features through the analysis of monogenic signal theory and monogenic binary algorithm. Then, the results of the facial expression recognition simulation experiment of monogenic based classical facial expression database, the Japanese Female Facial Expression (JAFFE) Database, and the results of traditional Local Binary Patterns-Sparse Representation-based Classification (LBP-SRC) residual fusion method were compared to illustrate the efficiency of the monogenic binary coding algorithm in the aspect of facial recognition and to provide a basis for the application of the monogenic signal theory in facial expression.

Povzetek: Opisana je nova varianta algoritma za hitro prepoznavanje izrazov na obrazu .

1 Introduction

Language is the most common way to convey information in daily life. However, the study shows that facial expressions contain more information than languages do and some cannot be easily recognized by humans themselves [1]. Thus, using computers to analyze facial expressions and to recognize important information has become a critical research topic in computer realm and the relative results have obtained a remarkable progress in the aspects of mental illness treatment, polygraph detection, distance education, human-computer interaction and so on. The rough operation flow of facial expression recognition is face detection and preprocessing, facial feature extraction, and facial expression classification [2], where face detection and preprocessing refer to the separation of the target face from the image; facial feature extraction refers to the extraction of some facial features that can reflect human emotions from the face, and facial expression classification is a classification of the extracted facial features as a specific expression. In the aspects of the extraction of facial features, there exist many mature technologies, including the extraction methods based on geometric features, overall statistical features, characteristic frequency rate, and motion features. Mollahosseini et al. [3] proposed a method to recognize facial expression with deep neural networks with good performance in blurred images. Dapogny et al. [4] proposed a method of facial recognition based on pairwise conditional random forests and its improvement in recognition was remarkable by learning random forests from different derivative characteristics in paired image. Lee et al. [5] proposed a new method of facial expression recognition based on sparse representation which could reduce the variance within class and emphasize to inquire

the facial expressions in face images at the same time. The monogenic binary coding algorithm is relatively more widely used among many facial expression recognition technologies, and many improved methods with high recognition rate have been derived over time. Yang et al. [6] proposed a monogenic binary pattern that combined monogenic orientations and monogenic amplitudes to apply on facial recognition realm, which obtained good results. Zhang et al. [7] proposed a monogenic binary algorithm based on local features which was relatively easier in calculation and improved the recognition accuracy further. In this study, facial expression recognition algorithm based on local features and monogenic binary coding was analyzed to build the facial expression recognition model through the deep understanding of the monogenic signal theory and the monogenic binary coding and the accuracy rate was examined by Japanese Female Facial Expression (JAFFE) Database, the classic facial expression database. Some suggestions about the methods of facial expression recognition were provided through the illustration of the wonderful features of monogenic binary coding in facial expression recognition realm, which was shown by the comparison of accuracy rate with the traditional facial recognition method.

2 Monogenic signal theory

Hilbert transform is one of the methods to describe signal analysis and complex envelope in the field of mathematics and signal processing and it can be roughly described as

the convolution of a signal $s(t)$ with $\frac{1}{\pi t}$ to get the signal $s'(t)$ [8].

The formula of Hilbert transform is:

$$s'(t) = \frac{1}{\pi t} * s(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{s(\tau)}{t - \tau} d\tau \quad (1)$$

Hilbert transform keeps the amplitude of each frequency component of the target signal and the phase exhibits a 90° shift.

Riesz transform is a two-dimensional extension of the Hilbert transform [9]. The monogenic signal based on Riesz transform can convert the one-dimensional signal into two-dimensional signal. The transformation form of Riesz transform in the spatial domain is:

$$f_R(z) = \begin{pmatrix} f_x(z) \\ f_y(z) \end{pmatrix} = \begin{pmatrix} \frac{x}{2\pi\|z\|^3} * f(z) \\ \frac{y}{2\pi\|z\|^3} * f(z) \end{pmatrix}, \quad (2)$$

where the convolution kernel of Riesz transform is $\left(\frac{x}{2\pi\|z\|^3}, \frac{y}{2\pi\|z\|^3}\right)$, and $*$ is convolution operator. The two-dimensional frequency response of the Riesz filter is $F_x = -\frac{jx}{\|x\|}, F_y = -\frac{jy}{\|y\|}$. The monogenic signal of a picture can be defined as the combination of the signal $f(z)$ and Riesz transform. The combination mode is:

$$f_s(z) = f(z) + (i, j)f_R(z) = f(z) + if_x(z) + jf_y(z). \quad (3)$$

The monogenic signal based on its features similarly as real and imaginary components can be orthogonally decomposed into local amplitude A , local phase φ , and local orientation θ . The three mentioned above relatively describe the local energy information, the local structure information and the geometric orientation information of the monogenic signal. The calculation method is:

$$\begin{cases} A = \sqrt{f^2 + f_x^2 + f_y^2} \\ \varphi = -\text{sgn}(f_x) \arctan\left(\frac{\sqrt{f_x^2 + f_y^2}}{f}\right) \\ \theta = \arctan\left(\frac{f_y}{f_x}\right) \end{cases} \quad (4)$$

3 Monogenic binary coding

It is necessary to analyze from several perspectives with different indexes to better describe the detailed information of an image. According to monogenic signal theory, an image can be extracted three specific pieces of information, including amplitude A , phase φ , and

orientation θ , then the local features of the image in details is described [10].

The feature images which have been decomposed can be named as monogenic binary coding amplitude image $JPG-A$, monogenic binary coding phase image $JPG-\varphi$, and monogenic binary coding orientation image $JPG-\theta$. The coding images which have been decomposed are all composed of two parts, intensity coding of monogenic local imaginary part and monogenic local variable coding. The coding formula after the combination is:

$$JPG-N(z_c) = [C'_x(z_c), C'_y(z_c), C_N(z_c)], N = A, \varphi, \theta, \quad (5)$$

where z_c refers to the central pixel of a selected area.

3.1 Intensity binary coding of monogenic local imaginary part

Since the Riesz transform is unsymmetrical in the monogenic signal, the features of the real part are suppressed, the representations of imaginary part are obvious, the decomposition effects are good, and the robustness is strong to factors such as illumination changes. In the image, the imaginary part after the monogenic binary coding can represent the intensity of local feature information of the central pixel. As for a pixel z_c , its intensity coding of monogenic local imaginary part is $[C'_x(z_c), C'_y(z_c)]$, and the coding rule is:

$$C'_n(z_c) = \begin{cases} 0, & f_n(z_c) > 0 \\ 1, & f_n(z_c) \leq 0 \end{cases}, n = x, y, \quad (6)$$

where f_x, f_y are the horizontal and vertical output of the imaginary part of Riesz transform represented by monogenic signal. The quadrant was expressed in the coordinate system in Figure 1.

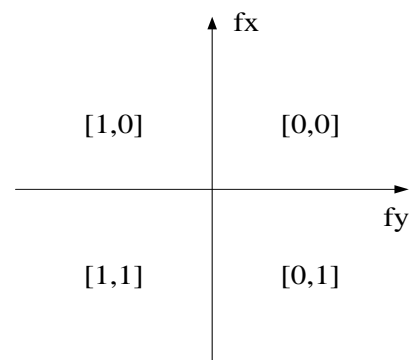


Figure 1: Quadrant coding of features of monogenic imaginary part in each place

3.2 Monogenic local variable binary coding

Local binary pattern (LBP) coding can be used in monogenic amplitude, local XOR pattern (LXP) can be used in monogenic orientation and monogenic phase and

the robustness of codes can be strengthened through blocking by the angles when generating the specific codes.

Local binary coding of monogenic amplitude usually uses round LBP operators. The working principle is to measure the local energy condition by comparing the amplitudes of the surrounding pixels and the central pixel. z_c is set as the central pixel and there are P proximal points in the surrounding area. The calculation method of amplitude binary coding of z_c is as follows:

$$C_A(z_c) = \sum_{i=0}^{P-1} s(A(i) - A(z_c)) \times 2^i, \quad (7)$$

where $A(i)$ is the amplitude value of the i -th proximal point, $A(z_c)$ is the proximal value of central pixel, and 2^i is the factor that converts binary system to decimal system. The function $s(x)$ is a sign function and its formula is as follows:

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \quad (8)$$

The XOR pattern coding of monogenic orientation and monogenic phase. Local orientation describes the major information of local structure, and local phase describes the texture information of the image. Both of them can be represented with the angles from 0 to 360, however, the special points will be too dispersed and unstable to extract local features, which refers to relatively worse robustness, in the situation where the groups are too much and trivial. Due to the strengthening of robustness of coding, it is necessary to divide those angles into several intervals to generate aggregation effect and highlight the features and steadiness of the centralized region.

In this study, the region, $[0, 360)$, was divided into 6 intervals. Different phases and orientation would be distributed to corresponding regions through categories and it was defaulted that the phases and orientation in the same region represented the same local features. The partition function $D(x)$ is shown as follows:

$$D(x) = \begin{cases} 1 & 0 \leq x < 60 \\ 2 & 60 \leq x < 120 \\ 3 & 120 \leq x < 180 \\ 4 & 180 \leq x < 240 \\ 5 & 240 \leq x < 300 \\ 6 & 300 \leq x \leq 360 \end{cases}. \quad (9)$$

The method of XOR pattern coding of central pixel z_c with N proximal points is as follows:

$$C_M(z_c) = \sum_{i=0}^{N-1} C_i^M \times 2^i, M = \varphi, \theta, \quad (10)$$

where C_i^M is defined as follows:

$$C_i^\varphi(z_c) = \begin{cases} 0, & D(\varphi(z_c)) = D(\varphi(i)) \\ 1, & D(\varphi(z_c)) \neq D(\varphi(i)) \end{cases}, \quad (11)$$

$$C_i^\theta(z_c) = \begin{cases} 0, & D(\theta(z_c)) = D(\theta(i)) \\ 1, & D(\theta(z_c)) \neq D(\theta(i)) \end{cases}, \quad (12)$$

where $C_i^\varphi(z_c)$ is the binary coding of local phase, $C_i^\theta(z_c)$ is the binary coding of local orientation, $\varphi(z_c)$ is the local phase of the center pixel z_c , $\varphi(i)$ is the phase of the i -th proximal point, $\theta(z_c)$ is the local orientation of the center pixel z_c , and $\theta(i)$ is the orientation of the i -th proximal point.

4 Facial expression recognition simulation experiment based on local features and monogenic binary coding

4.1 Experimental preparation

The experiment was run on a desktop computer with a Windows 10 32-bit system, quad-core 3.30 GHz, and 4 GB memory.

This experiment is a non-specific facial expression experiment, using JAFFE Database which is a classic facial expression database. The database consists of 7 expressions, each with 3 to 4 sample images taken from 10 Japanese women, and there were 213 facial expression images. The seven types of expressions are sadness, joy, anger, nausea, surprise, fear and neutral expression [11].

4.2 Experimental procedure

Test set and training set were selected firstly. The experiment was divided into 10 rounds. In every round, 9 out of 10 females and about 190 facial expressions would be selected as the training set, and the remaining one about 20 facial expressions was as the test set. After the end of the 10th round, the average value of the recognition rates was calculated as the result of this experiment. The extracted expressions of a woman in the library are in Figure 2.



Figure 2: Seven expression examples in JAFFE

In order to ensure the validity of the recognition, it was firstly necessary to perform preprocessing. A uniform cropping template was taken to remove redundant information such as hair and neck, and then the processed



Figure 3: The example of preprocessed results.

images were obtained after being equalized and scaled. The examples are shown in Figure 3.

The feature graphs of monogenic binary coding of each image could be obtained with the method introduced in this study and the feature graphs $JPG - A$, $JPG - \varphi$, and $JPG - \theta$ could be generated correspondingly with amplitude A , phase φ , and orientation θ . Eight proximal points in the proximal region could be obtained with the radius of 2 in the LBP calculation. The coding values of LBP coding and LXP coding were both in the range from 0 to 255; the monogenic local intensity coding was non-successive, from 0 to 1024.

A histogram was established with the three features to further construct the feature histogram of the partial block. The images were divided into 15 non-overlapping blocks provided that the faces were standardized. Then every block were divided again into 3×3 sub-blocks. The feature graphs of the sub-blocks in the three dimensions were put in series to obtain the histogram of the sub-blocks. The feature histograms of every partial block could be obtained as a proof after being connected.

Fisher linear discriminant analysis was used to reduce dimension of the training sets block by block [12]. The original image had a large amount of information and a large number of dimensions. If the identification was directly recognized, a "dimension disaster" might occur, thus it was necessary to perform dimension reduction at the beginning. Firstly, the linear discriminant analysis was used to learn the feature histogram matrix of each local block, and the best projection vector was obtained. Then, it was convolved with the histogram to obtain the reduced-dimensional histogram.

The cosine distance method was used to calculate the similarity between the test set and the training set in each local block, and the recognition was accumulated at the final [13]. The facial recognition of monogenic binary coding after fusion was supposed as $JPG - Com$, and the results of weighting fusion is as follows:

$$S_{JPG-Com} = 0.27S_{JPG-A} + 0.27S_{JPG-\varphi} + 0.46S_{JPG-\theta} \quad (13)$$

The LBP-SRC residual fusion algorithm with high recognition rate in traditional face recognition technology was selected [14], the results are modeled and calculated in the method described in this study, and the calculation results were obtained after being compared with the monogenic binary coding method.

4.3 Experimental results and analysis

The comparison results of the recognition rates are showed in Table 1.

Type of recognition algorithm	LBP-SRC residual fusion	$JPG - A$	$JPG - \varphi$	$JPG - \theta$	$JPG - Com$
The success rate of recognition /%	71.15	66.37	68.92	70.38	78.71

Table 1: Comparison results of algorithm recognition rate.

It could be easily seen that the traditional LBP-SRC residual fusion method had a relatively good recognition rate, which reached 71.15%; although the three separate monogenic binary coding calculation methods also had relatively higher facial expression recognition rate, their rates were lower than those of the traditional LBP-SRC residual fusion method.

The algorithm based on local features and monogenic binary coding integrated the recognition characteristics of the three features, and reduced the volatility and limitation of recognition. Therefore, the highest facial expression recognition rate obtained was 78.71%.

The corresponding facial expression recognition rate based on different expression categories is shown in the Figure 4.

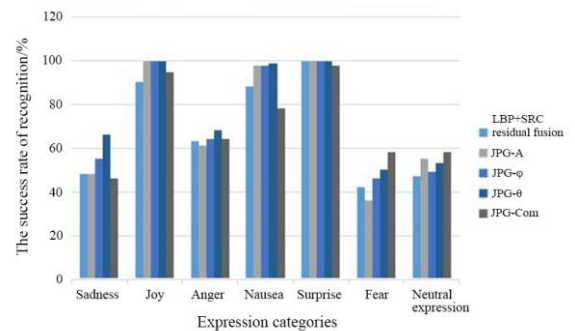


Figure 4: Comparison of five algorithms in the recognition rates of seven categories of expressions.

It could be seen that the recognition rates of the five modes were different in different types of expressions. The total recognition rate seemed to be higher in the expressions of joy, nausea, and surprise. The most easily recognized expression was surprise. Sadness, anger, fear, and neutral expression were more difficult to recognize because of the small degree of facial muscle changes. The most difficult expression to recognize was fear.

After comparing the recognition rates of different algorithms in the same expressions, it was known that the recognition rates of $JPG - Com$ was higher than those of $JPG - A$, $JPG - \varphi$, and $JPG - \theta$ except the neutral expression, and the highest recognition rate of $JPG - Com$ was in the three expressions of joy, nausea, and surprise. After comparing of $JPG - Com$ and LBP-SRC residual fusion, it was known that the recognition rates of $JPG - Com$ in the expressions of sadness, joy, anger, nausea, and surprise were higher than those of LBP-SRC

residual fusion, but the recognition rates in the expressions of fear and neutral expression were lower than those of LBP-SRC residual fusion.

Based on the above research and considerations, it could be said that the algorithm based on local features and monogenic binary coding was a more effective facial expression recognition method with higher recognition efficiency.

5 Conclusion

Nowadays, facial recognition technology is widely used. In this study, facial expression recognition based on local features and monogenic binary coding algorithm was analyzed to illustrate the principles of monogenic signal theory and monogenic binary coding, and then the corresponding facial expression recognition model of the monogenic binary coding algorithm was constructed according to the theory, which was then applied to the JAFFE Database samples. Compared to the traditional LBP-SRC residual fusion method, the effectiveness of the monogenic binary coding algorithm in the face recognition field was shown, providing some references for related research.

References

- [1] Kim JG, Ko JG, Lee SJ (2016) Demo:MilliCat: Real-Time Autonomous Image Suggestion for Mobile Messaging. *International Conference on Mobile Systems*. <https://doi.org/10.1145/2938559.2938585>.
- [2] Kim BK, Roh J, Dong SY, Lee SY. (2016) Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *Journal on Multimodal User Interfaces*, 10(2), pp. 173-189. <https://doi.org/10.1007/s12193-015-0209-0>.
- [3] Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Lake Placid, NY, USA. <https://doi.org/10.1109/WACV.2016.7477450>.
- [4] Dapogny A, Bailly K, Dubuisson S (2016) Pairwise Conditional Random Forests for Facial Expression Recognition. *IEEE International Conference on Computer Vision*, IEEE, Santiago, Chile. <https://doi.org/10.1109/ICCV.2015.431>.
- [5] Lee SH, Ro YM (2017) Intra-Class Variation Reduction Using Training Expression Images for Sparse Representation Based Facial Expression Recognition. *IEEE Transactions on Affective Computing*, 5(3), pp. 340-351. <https://doi.org/10.1109/TAFFC.2014.2346515>.
- [6] Yang M, Zhang L, Zhang L, Zhang D (2010) Monogenic Binary Pattern (MBP): A Novel Feature Extraction and Representation Model for Face Recognition. *International Conference on Pattern Recognition*, IEEE, Istanbul, Turkey. <https://doi.org/10.1109/ICPR.2010.657>.
- [7] Yang M, Zhang L, Shiu CK, et al. (2012) Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition. *IEEE Transactions on Information Forensics and Security*, 7(6), pp. 1738-1751. <https://doi.org/10.1109/TIFS.2012.2217332>.
- [8] Culiuc A, Kesler R, Lacey MT (2016) Sparse Bounds for the Discrete Cubic Hilbert Transform. <https://doi.org/10.2140/apde.2019.12.1259>.
- [9] Lunyov AA, Malamud MM (2016) On the Riesz basis property of root vectors system for 2×2 Dirac type operators. *Journal of Mathematical Analysis & Applications*, 441(1), pp. 57-103. <https://doi.org/10.1016/j.jmaa.2016.03.085>.
- [10] Wang Z, Sun X, Sun L (2016) Face Recognition Using Monogenic Signal Feature and Discriminative Manifold Preserving Embedding. *Journal of Computational and Theoretical Nanoscience*, 13(4):2537-2550. <https://doi.org/10.1166/jctn.2016.4615>.
- [11] Saaidia M, Zerme N, Ramdani M (2016) Multiple Image Characterization Techniques for Enhanced Facial Expression Recognition. *Intelligent Systems Technologies and Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-319-23036-8_43.
- [12] Mahdianpari M, Salehi B, Mohammadimanesh F, Brisco B, Mahdavi S, Amani M, Granger JE. (2018) Fisher Linear Discriminant Analysis of coherency matrix for wetland classification using PolSAR imagery. *Remote Sensing of Environment*, 206, pp. 300-317. <https://doi.org/10.1016/j.rse.2017.11.005>.
- [13] Wang X, Yang S, Zhao Y, Wang Y (2018) Lithology identification using an optimized KNN clustering method based on entropy-weighted cosine distance in Mesozoic strata of Gaoqing field, Jiyang depression. *Journal of Petroleum Science and Engineering*, 166, pp. 157-174. <https://doi.org/10.1016/j.petrol.2018.03.034>.
- [14] Ameer B, Masmoudi S, Derbel AG, Hamida AB (2016) Fusing Gabor and LBP feature sets for KNN and SRC-based face recognition. *International Conference on Advanced Technologies for Signal & Image Processing*, Monastir, Tunisia. <https://doi.org/10.1109/ATSIP.2016.7523134>.

