

Feature extraction of smiley facial expression based on AU sequence optical flow

Jin Yan^{1, a)}, Jin Wang^{2, b)} and Qing Zhu^{2, c)}

¹School of Beijing University of Technology, Beijing 100124, China.

²School of Beijing University of Technology, Beijing 100124, China.

³School of Beijing University of Technology, Beijing 100124, China.

a)^{yanjin728@163.com}, b)^{jijnwang@bjut.edu.cn}, c)^{ccgzq@bjut.edu.cn}

Keywords: Smile Snapshot, AU sequence, Optical flow, Signature Sequence

Abstract. Facial expression recognition is widely used in linguistics, medical care, service and so on. . But still there exists the problem of accuracy.. To solve this problem, , this paper presents an extraction method based on the flow characteristics of facial motion unit for the facial expression recognition. The facial action coding system (FACS) is first applied to extract facial motion unit, and then optional flow method for these areas is used to extract eigenvalue to constitute the signature sequence, finally categorization based on random forests is applied to classify and recognize the smiley facial expression. Experimental results validate the effectiveness of the proposed scheme.

INTRODUCTION AND RELATED WORKS

Introduction

Emotion recognition is a technique for classifying deformation angle from facial movement or facial features by studying the visual information of images. In the 1970s, Suwa and Sugie made an initial attempt at the recognition of facial expression based on dynamic image sequences. They selected 20 representative feature points. In order to obtain the regularity of the motion of these 20 feature points, they tracked a face video animation and carried out expression recognition research work by comparing and analyzing the law of motion change with the motion change model of various expressions which have been established in advance. Kenji Mase et al. of Japan ART Research Institute proposed a facial expression moving unit tracking method based on studying optical flow method, and then worked on facial expression recognition based on dynamic image sequence. In addition, Cohn and Kanade et al. used the optical flow method to identify the subtle changes of facial expressions, and automatically tracked and obtained facial expression dynamic characteristics through optical flow grading calculation method, and carried on the experiment about the eyebrows and mouth area and Mixed action unit. The facial recognition method of Yacoob and Davis is based on Facial Action Coding System (FACS). In China, Jinhui of Harbin Institute of Technology and Gao Wen of Chinese Academy of Sciences have proposed a facial mixed expression recognition system, the overall recognition rate of which has reached 96.9%. Peng Yang, Qingshan Liu proposed a method based on dynamic feature coding for facial expression recognition, and applied this method to face recognition based on video. Yang applied the Cohn-Kanade expression library to experiment with the proposed algorithm, and compared it with the facial expression recognition method under static images. The results show that the recognition result of dynamic image sequence is better than static image recognition results. Jia li Gou used the modified feature-point matching algorithm of Speeded Up Robust Features (SURF) to extract the moving regions and carried out experiments. In order to solve the problem of facial expression recognition in the context of single video dynamic change, Yu Jun, Zengfu Wang, Rui Li proposed a synchronous face tracking and facial expression recognition algorithm.

In 1971, two American psychologists, Ekman and Frisen, divided the expression into six categories, respectively, anger, happiness, sadness, surprise, disgust and fear [1]. FACS [2], which

is associated with human expression, was later used in the study. Among the facial expressions, smile is one of the most important ones used for interpersonal and man-machine communication. Especially the use of smile recognition in life, like patient's mood analysis, face snapshot function in intelligent mobile phone, etc. But there are still many difficulties and challenges in the field of smile snapshot, like the accuracy and timeliness of smile snapshot. Therefore, smile snapshot of dynamic series is a hot issue we need to research and rescue.

FACS

FACS, is also known as the Facial Action Coding System or Facial Behavior Coding System, which depicts the corresponding relation between different facial muscle movements and different emotions. Through the summary and analysis of different parts of the facial action, the researcher obtained 44 independent and correlative motor units AU (Action Unite), and used different coding numbers to represent these facial movements, such as AU1, AU2, AU3 and so on. These independent AU control region also communicate with each other during the face morphing time. It can synthesize nearly 10000 kinds of expressions through the free combination of coding, which including over 3000 emotional sense of expression [4]. Nowadays, this coding system has been widely used in facial expression recognition and synthesis, as well as the relative calculation.

Analysis of Smile Expression Features

According to the facial motion coding system, the smiley facial expression is composed of motor unit AU6 and AU12. AU6 shows the contraction of the orbicularis oculi muscle, the performance is that cheek upwards and eye wrinkles appear. AU12 shows the contraction of the zygomatic major muscle, the performance is that mouth stretches upwards and to either side. AU12 pull on both sides of the mouth tilt upward, while pushing up the box under the triangle area accumulation. The meaning of action is very simple, is representative of the laugh (here is not really laugh and smirk discussion). When the AU appears, it can judge that the litigant is laughing. When joined AU6, it can almost completely determine that the litigant is in a happy mood. As show in Fig. 1, it is a three-frame diagram of the sequence of smiling images. (a) is a neutral expression, and (c) is the final expression. In this process, we can observe, this time there are facial features: lips open, mouth upturned, while his eyes slightly closed.



FIGURE 1. three-frame smiling images

Calculation of optical flow

Optical flow method is an important method of motion image analysis at present. It is a method in the gray mode, which uses the change of Pixel Grayscale Value in time series in image sequence and the correlation between adjacent frames to find out the existed corresponding relation between the previous frame image and the current frame image, in order to calculate the motion information of the objects between adjacent frames.

Suppose that at time t , Pixel coordinates in the image is (x, y) , the gray level is $I(x, y, t)$. After the time Δt , this point moves to $(x + \Delta x, y + \Delta y, t + \Delta t)$, its gray level is $I(x + \Delta x, y + \Delta y, t + \Delta t)$. Because it is the same point at two different moments, according to the assumption of gray-scale conservation, it has

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (1)$$

Using Taylor formula to simplify the right side of equation (1), we can obtain the basic equation of optical flow:

$$\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} = 0 \quad (2)$$

In the formula $u = \frac{dx}{dt}$, $v = \frac{dy}{dt}$, are the moving velocities of (x, y) in the x and y directions during the time interval $\Delta t = t_k - t_{k-1}$.

In addition, after observation, we can know that there are two variables in equation (2). So there must be additional constraints to solve the equation. The two most commonly used methods are Horn-Shunck and Lucas-Kanade. This paper selects the Horn-Schunck algorithm.

Horn-Schunk

Horn-Schunck algorithm suggests adding smoothness constraint of optical flow. The core idea: Each optional flow on image is in contact. Option flow in the entre image should be smooth changes, so we introduce the smoothness constraints to minimize the smooth constraint.

Suppose that the smoothness constraint term is minimized:

$$E_x = \iint (u_x^2 + u_y^2 + v_x^2 + v_y^2) dx dy \quad (3)$$

Combine with the basic equations to require the minimization:

$$E_c = \iint (I_x u + I_y v + I_t)^2 dx dy \quad (4)$$

The resulting optical flow should be satisfied:

$$\min \iint |(u_x^2 + u_y^2 + v_x^2 + v_y^2) + \lambda(I_x u + I_y v + I_t)^2| dx dy \quad (5)$$

Compose the gradient equation (2) and smoothing constraint equation (5) into equation set and discretize, the recursive formation of the optical flow can be obtained as shown in the following formula (6)

$$\begin{cases} u_{kl}^{n+1} = u_{kl}^{-n} - I_x \frac{I_x u_{kl}^{-n} + I_y v_{kl}^{-n} + I_t}{1 + \lambda(I_x^2 + I_y^2)} \\ v_{kl}^{n+1} = v_{kl}^{-n} - I_y \frac{I_x u_{kl}^{-n} + I_y v_{kl}^{-n} + I_t}{1 + \lambda(I_x^2 + I_y^2)} \end{cases} \quad (6)$$

PROPOSED METHOD

Image acquisition

The Cohn-Kanade of CMU is a facial expression database including 2000 gray - scale image sequence of over 200 people. Each person in the database contains six basic expressions, each expression sequence changes from the neutral to the final expression, each second is about 20-30 fps, and the image size is 640 * 480 pixels. This database in a facial expression database is based on motion unit code. According to the facial action coding system, it divides the facial action into 44 kinds, and collects most of the basic AU sequences of human faces.

Before extracting the feature of facial expression, we need to detect and preprocess the face image. In this study, the face detection stage is realized by combining Haar feature and Adaboost algorithm. In order to obtain one person's face image, we need to normalize its size. We use a specific coordinate method that fixes the coordinates of the eye to the output image, use the principle of affine transformation, and then obtain a transformed new image. As it is shown in Figure 2.



FIGURE 2. Normalized image sample

Feature region partitioning

Unlike superficial features of face images such as gray scale, texture, etc., facial expression mixture features can describe the expression more accurately based on facial motion unit. Most previous work has divided the image into many sub-blocks and extracted facial motion unit information from them. In this article, we will extract the image sub-block from the eye regions and mouth regions of face image, and then compose characteristic region, for subsequent use in conjunction with the optical flow method.

According to that the gray gradient of the eye is the largest among the five parts of the human face [7], we use the proximate point grey level difference formula to calculate the gray scale gradient difference, and set the threshold to separate eye area.

Since we locate the eyes of normalized image, thus, add face geometry knowledge in the literature method to improve the location accuracy. After the precise positioning of the human eye, we use a positioning method about oral horizontal center line to position oral area. As shown in Figure 3.



FIGURE 3. the expression of eyes and mouth after the facial expression

After the initial positioning of the human eye and mouth, it is necessary to extract and divide the region of local facial motor unit in the facial expression map; the definite means are as follows:

Supposed that the image size is $m*m$. First, see human eye horizontal center line as horizontal center, $m/8$ as the horizontal movement step of the child window, $m/8$ as the home position in the vertical direction, and then extract five sub-windows of size $m/4*m/4$ as upper half face facial motion unit sub-block set (that is the areas of eyebrows and eyes). Similarly, see oral horizontal center line as horizontal center, $m/8$ as the horizontal movement step of the child window, $m/8$ as the home position in the vertical direction, and then extract five sub-windows of size $m/4*m/4$ as lower half face facial motion unit sub-block set (that is the areas of mouth). As shown in Figure 4.

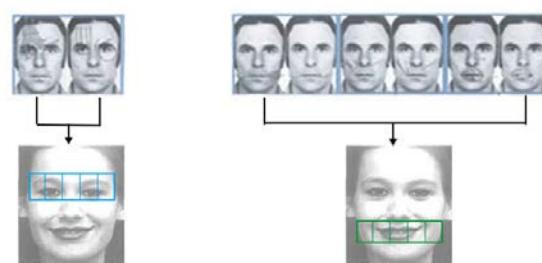


FIGURE 4. the relationship between local sub-block and facial motion

The improvement of Horn-Schunck optical flow calculation method

The Horn-Schunck computation of optical flow fields is based on the transient conservation of gray-scale of moving objects. When the light changes or when there is deformation of moving objects, illumination calculation will result in an error, and the motion information of the face part cannot be extracted correctly. For this problem, relevant scholars have proposed an improved method, the idea is to incorporate the characteristics of the non-rigid fluid motion based on the optical flow basic extension algorithm of the motion continuity equation, to use the divergence and rotation spline functions as an additional constraint factor term [5]. Compared with the traditional Horn-Schunck method, the optical flow constraint equation for non-rigid objects adds $I_{div}(U)$ term, and introduces $div-curl$ spline function. τ, φ are the smoothing parameters of walking and curl constraint respectively. If τ or φ is larger, the contribution of the corresponding divergence or curvature spline constraint to the optical flow computation becomes larger, and vice versa. Reference the method [5] in literature, after determining the parameters, we can obtain the optical flow data.

K-L transformation on optical flow data

Extract the optical flow data dimension of dynamic characteristics directly is not suitable for classifying directly. In this paper, the principal component analysis [11] can be used to dimension reduce the optical flow data. Principal component analysis (PCA) is also referred to as K-L transformation, it is collection and transformation of database, lets the new data have the given statistical properties, which is widely used in the aspects of data classification and compression, etc.

Optical flow dynamic feature extraction of AU sequences

Before the computation of optical flow, we divide the moving units into regions according to the method described in section 4.2 above, the advantage of this approach is that:

- a) There is no need to optical flow calculate the entire face, greatly reduce the amount of computation
- b) Analysis for specific areas, which can remove other redundant information in the face
- c) Compared with the whole face processing, data latitude is significantly reduced after the regional division

In this paper, we choose the normalized facial expression image as the initial object, extract the upper half face 5 feature region subsets and the lower half face 5 feature region subset block AU image in the initial object as target sequence. Derive the sequence characteristic value includes two parts; derive PCA substrate and combine AU eigenvalue.

Obtain the space base of PCA dimensionality reduction

Firstly, create a sample space, obtain the base of PCA feature projection, and obtain the final conductivity. Then, Divide each face image sequence into regions, and obtain the corresponding AU sequences. Calculate the optical flow field of each image in the AU sequence to obtain the u and v components. Due to different image sizes in different regions, put the optical flow data in a sub-space in the same area, that is, crate a different data subspace for the upper half of the face and the lower half of the face, denoted by Q_u, Q_d respectively. The PCA algorithm is used to calculate the base for the u and v components in each subspace data Σ_Q^u, Σ_Q^v . . Then each subspace contains u, v components of the substrate, that is $\Sigma_{Q_u}^u, \Sigma_{Q_u}^v, \Sigma_{Q_d}^u, \Sigma_{Q_d}^v$.

Combining AU eigenvalues

- a) Given an image sequence of testing the AU unit. After obtaining the optical flow field of the i-th frame image, correspond it to the base space according to sub region, and obtain the projection conductivity of the component in this frame.

$$\begin{cases} p_i(\Sigma_Q^u) = [P_{i1}^u, P_{i2}^u, P_{i3}^u, \dots, P_{in}^u] \\ p_i(\Sigma_Q^v) = [P_{i1}^v, P_{i2}^v, P_{i3}^v, \dots, P_{in}^v] \end{cases}$$

- b) Respectively, take the first M (M is usually small) in projection conductivity $P(\Sigma_Q^u)$, $P(\Sigma_Q^v)$, and link the conductivity end to end to form the eigenvalues of the i-th ($i = 1, 2, \dots, L$) frame image in this AU sequence, then:

$$T_i = [P_{i1}^u, P_{i2}^u, P_{i3}^u, \dots, P_{iM}^u, P_{i1}^v, P_{i2}^v, P_{i3}^v, \dots, P_{iM}^v]$$

- c) Link the eigenvalues of all L frames end to end in the test AU sequence to form the final eigenvalue O of this sequence, so the AU sequence is characterized by $O = [T_1, T_2, T_3, \dots, T_L]$. O is row vector of $2 * M * L$ dimension.
- d) For K different test sequence image composed by different AU sequences, according to the above steps, the sequence of the characteristic values can be constructed as follows:

$$Seq = [O_1, O_2, \dots, O_K] = [T_{11}, T_{12}, \dots, T_{1L}, T_{21}, T_{22}, \dots, T_{2L}, \dots, T_{k1}, T_{k2}, \dots, T_{KL}]$$

From the above analysis, we can get two types of eigenvalue sequences. For convenience, Seq_u is defined as on behalf of the upper half face and Seq_d is defined as the lower half face.

EXPERIMENTAL RESULTS

The AU instances of the mouth

In this paper, different types of AU images from 30 testers were collected. All images in the samples were normalized to a size of $64 * 64$ pixels. The size of the extracted facial motion sub-block is $16 * 16$. Take the lower half of the face mouth region for example, the mouth area includes 10 AU, AU sequence length is $L=3$, which constitute the PCA sample space of 1020 frame images. According to the PCA dimensionality reduction substrate ($f=1000$), it can obtain the PCA substrate of optical flow data u, v component in mouth area respectively.

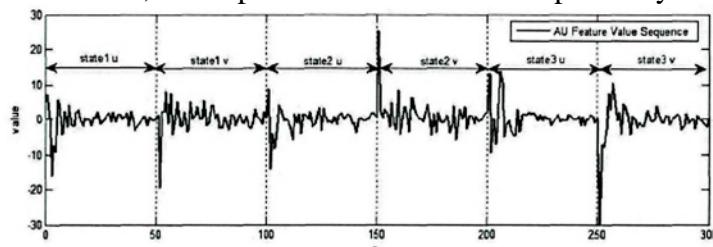


Figure 5. the PCA substrate of optical flow data u, v component in mouth area

The eigenvalues corresponding to the AU instances of the mouth are given. The u, v components of each data is taken as the first 60 ($M = 50$) projection coefficients and represented in one-dimension. The whole feature extraction process ends with the acquisition of the eigenvalue sequence, and the sequence of eigenvalues is important for subsequent classification. The observed value are extracted from the sequence of eigenvalues in a subsequent process, is that an observed value corresponding to a single state of the eigenvectors a single state of the eigenvectors $O = T_i (i = 1, 2, \dots, L)$ in the eigenvalues of the AU dimension $O =$ a single state of the eigenvectors $T_i (i = 1, 2, \dots, L)$.

Facial expression classification

The use of classifiers in recognition of smiling faces has two main roles: training and recognition. In this paper, the classification and recognition module adopts random forest classification method. The techniques include bagging random selection method, classification regression tree (CART) algorithm, Gini coefficient minimum principle, out-of-pocket estimation value and so on. The random forest was Forests-RI. Two stochastic algorithms construct the random forest used in this paper: Bagging (Bagging) and building method based on the input.

As a result of acquisition method of random forest algorithm's data is: training and selecting sample simultaneously in a database, and distinguish the training of training samples and the identification of test samples. So, all samples are stored in the same folder. This paper uses CMU's Cohn-Kanade expression library for experiments. In order to obtain the classifier with better recognition effect, the number of positive and negative samples should be balanced. Selecting 20

individuals from the data and selecting 10 individuals from the text. And using further 10 individual's facial expression image sequences in the text. A total of 250 training images were obtained, including 10 neutral expression maps, 241 facial expressions. The sample data is assigned as the following table1:

TABLE 1. Sample data.

Total	Neutrality	Smile	Other
250	10	115	125

The test results are shown in the following table2:

TABLE 2. the test results

Total	Total correct	recognition rate
241	239	99.17%

CONCLUSION

The experimental results show that the optical flow method can track the feature area of facial expression, which is fast and efficient. The extracted facial expression is relatively accurate and stable. Relative to other facial expressions, the expression of pleasure is more obvious, the corresponding optical flow is also more significant. Facial motion units are extracted by face facial expression coding system (FACS), and then the optical flow method is used to extract the eigenvalues of these regions, which are used to identify the facial expressions. Compared with the facial expression recognition system using static face images, the expression sequences contain more expression information. Therefore, the recognition accuracy and reliability are improved. The further work is to improve the feature extraction method, and consider whether local features can be combined with optical flow characteristics. So that, more accurate recognition results will be obtained.

REFERENCES

- [1] Ekman P, Friesen W V. Constants across cultures in the face and emotion.[J]. *Journal of Personality & Social Psychology*, 1971, 17(2):124-9.
- [2] Ekman P, Rosenberg E. What the face reveals : basic and applied studies of spontaneous expression using the facial action coding system (FACS)[J]. Oxford University Press, 1997, 68(1):83-96.
- [3] Wang Zhiliang, Chen Fengjun, Xue Weimin. "A Survey of Facial Expression Recognition" [J]. *Computer Applications and Software*, 2003, 20(12):63-66.
- [4] Chen Xuxu, "Facial Image Expression Generation based on Candide - 3 Model" [D]. Shandong University of Finance and Economics, 2014.
- [5] Yang Guoliang, Wang Zhiliang, Wang Guojiang and so on. Et al. Facial expression recognition based on non-rigid motion optical flow algorithm [J], *Computer Science*, 2007, 34 (3): 213-215.
- [6] Huang Xia. An Image Data Indexing Algorithm Based on Ontology and Singular Value Decomposition[J]. *Journal of Electronics*, 2014, 2 (42): 287-291.
- [7] Wang Yinghui. Face Recognition: Principles, Methods and Techniques [M]. Science Press, 2010.
- [8] Bartlett M S, Littlewort G C, Frank M G, et al. Automatic Recognition of Facial Actions in Spontaneous Expressions[J]. *Journal of Multimedia*, 2006, 1(6):360 - 365.
- [9] Ouyang Yan, Sang Nong. Facial expression unit recognition based on facial expression [C] // National Conference on Image and Graphics. 2010: 38-43.
- [10] Ren Yan. Face AU combination recognition based on key point LBP feature extraction [J]. *Information technology research*, 2012 (2): 44-47.
- [11] Pang Cheng. Research on Subspace Reduction Methods in Face Recognition[D]. Yangzhou University, 2015.