Extraction of Local Expression Features through Combination with DT_CWT and DCT Based on ASM

Ping Luo, Xianfei Li [a], Liangxue Huang, Qiang Ma

College of automation, Chongqing University of Posts and Telecommunications, Chongqing, China

[a]lixianfei@yahoo.com

Abstract

Aimed at imprecise key feature points of facial expression obtained through location of Active Shape Model (ASM), which caused position displacement in Region of Interest (ROI), and resulted in problems such as blurring and unobvious local texture features of extracted feature points, this paper proposes an extraction method for expression texture features through combination with Dual-tree _ Complex Wavelet Transform (DT_CWT) and Discrete Cosine Transform (DCT). Firstly, two-dimensional filtering shall be conducted for key feature points after location by using 4-level DT_CWT, to prevent imprecise location which will lead to position displacement in ROI. Secondly, local texture features of expression shall be extracted by using DCT. Thirdly, final feature vector shall be formed through selection of parameters of which DCT coefficient is larger by Zig-Zag method. The experiment shows that this method can improve the blurring extraction of local texture features caused by imprecise location for key feature points well, and increase the recognition rate of facial expressions.

Keywords

Active Shape Model; Dual-tree _ Complex Wavelet Transform; Discrete Cosine Transform; expression recognition.

1. Introduction

Facial expression is the most important way of transmission of emotional information and communication. Along with the advance of computer and image processing technology, facial expression recognition has important research significance in human-computer interaction, image processing, emotional computing, and machine vision [1].

ASM was firstly proposed by Coots [2] et al., which can cover the subspace of facial geometric shape well, get global features of facial expression based on distance, and better supplement description of fine features in key expression feature region through extraction of local texture features of located feature points. Thus, ASM has been widely used for the extraction of facial expression features. However, extraction of local texture features based on ASM will be easily affected by position displacement of located feature points, resulting in blurring and unobvious extracted local texture features [3], and leading to unsatisfactory recognition rate of similar and subtle expressions, which has great influence on timeliness and feeling of man-machine interaction. Thus, how to extract clearer and more obvious ASM local texture features has become a research focus and difficulty in the area of facial expression recognition [4]. In 2014, Zhu Shaoping et al. adopted ASM model to locate feature points, and got expression features integrated by facial geometrical feature and local texture feature. This method has obtained better effect on the extraction of global feature and local texture feature [5]. However, location of this method shall be manually calibrated, which is very tedious. This problem can be solved through file automatic calibration of facial markers, but images in face database and of facial expressions adopted in this method have some difference, leading to position displacement of key feature points after calibration, i.e. displacement of ROI area, which will affect the following accurate extraction of local texture features [6].

The translation invariance of DT_CWT [7] will protect signal from being affected by position displacement through filter with orthogonal construction, total reconstruction and Hilbert Transform to each other. Based on this, the paper proposes expression feature extraction method of DT_CWT+DCT.

2. Location of ASM Key Feature Point

58 key feature points have been calibrated as shown in Figure 1 by using images from IMM face database and calibration point documents in this paper. Feature point set of target contour has been extracted, which forms a training set; then make model approaches to actual target contour step by step through the training set, so that to achieve the goal of accurate location of key feature points. Key feature points located by ASM have been tested through collection of some images on facial expressions in this paper. Location effect of a participant is as shown in Figure 1.

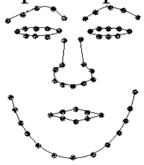
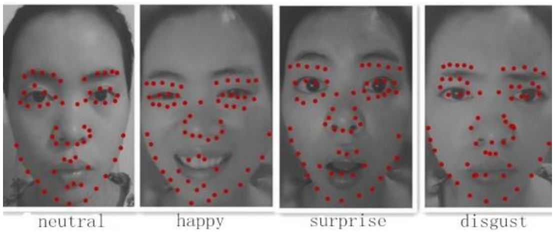


Figure 1 58 Feature Points



Feature 2 Location Effects of Feature Points

Figure 2 shows location effects of key feature points of neutral, happy, surprise and disgust successively. In the figure, accurate location of the neutral and unsatisfactory location effect of the other expressions can been presented. Different degree of deviations occurs in the locations of feature points of eyebrow, mouth and eyes, which leads to position displacement of feature points, and will affect the accurate extraction of following local texture features. Since IMM library is not an expression database, which does not include each facial expression action, it cannot adapt to changes of expression in the search of location. Thus, displacement of key feature points occurs after location.

3. Extraction of Local Expression Features through Combination with DT_CWT and DCT Based on ASM

After positioning key feature points of expression, local texture features should be extracted and in this paper, we adopt DCT method. Due to deviation in positioning feature points, the extracted features are not exact. Therefore, this paper introduces DT_CWT for improvement and the overall algorithm process is as follows in figure 3. Overall process of facial expression recognition is indicated in figure 4 and the paper is to improve extraction of ASM local features.

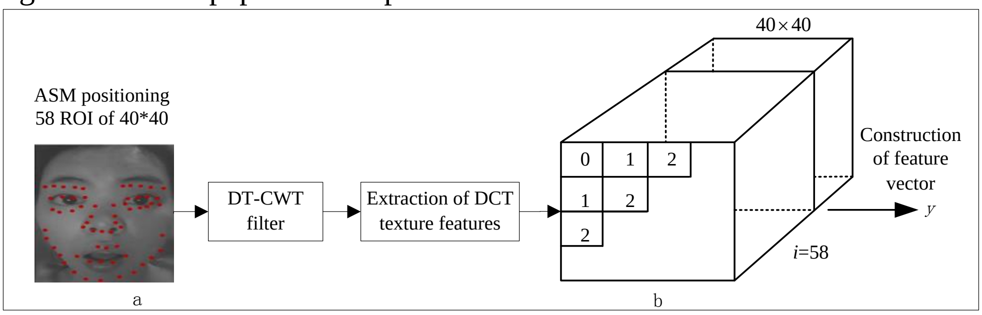


Figure 3 Process of extraction of ASM local expression textural features improved by DT_CWT+DCT

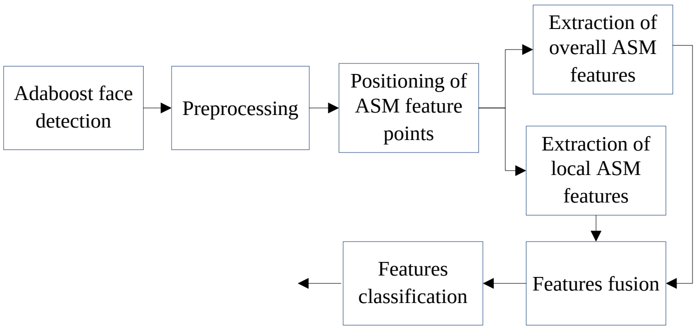


Figure 4 Overall flow diagram of facial expression recognition

First of all, a two-dimensional filtering is conducted to the region of interest of positioned key feature points by 4-level DT_CWT, and then textural features of feature points are extracted by DCT. Image features undergone DT_CWT filtering include 6 directions in each level, i.e. image filtering of each feature point produces 4 dimensions with 6-directional sub-bands on each dimension and the image features include 24 sub-band matrix in total. Each sub-band matrix could be transformed to real matrix (i.e. amplitude matrix of the sub-band) by computing amplitude value of complex coefficient. Each amplitude matrix is expanded in columns to form a column vector which shall be indicated by $V_{l,\theta}$, in which l and θ representing decomposition layer-number and direction parameter respectively (value range: $\theta \in \{+15^{\circ}, +45^{\circ}, +75^{\circ}, -75^{\circ}, -45^{\circ}, -15^{\circ}\}$; $l \in \{1, \cdots, 4\}$). Feature vector X of each feature point image undergone DT_CWT is structured and formed by column vectors corresponding to the 24 sub-bands, which could be indicated as follows:

$$\begin{aligned} X = \{ & V_{1,+15^{\circ}}^{\mathrm{T}}, V_{1,+45^{\circ}}^{\mathrm{T}}, V_{1,+75^{\circ}}^{\mathrm{T}}, V_{1,-15^{\circ}}^{\mathrm{T}}, \\ & V_{1,-45^{\circ}}^{\mathrm{T}}, V_{1,-75^{\circ}}^{\mathrm{T}}, V_{2,+15^{\circ}}^{\mathrm{T}}, \cdots, V_{4,-75^{\circ}}^{\mathrm{T}} \}^{T} \end{aligned} \quad (1)$$

In which the superscript T represents transposition operation. In formula (1), the dimension of feature vector X is equal to summation of the number of coefficients produced on 6 directions of each of the four levels of DT_CWT filters which leads to a large space dimensionality of feature vector X resulting in great burden in computation and recognition speed. Therefore, the dimensionality of feature vector should be reduced after DT_CWT filtering and DCT transforming of feature vector X. DCT could reflect spatial correlation. For flat regions without texture, cosine transform only involves with average grey degree i.e. (0，0) components. For regions with rough texture, images have spatial correlation with a large span distance and $|C(u,v)|^2$ would obtain a large value at low frequency component i.e. $\sqrt{u^2+v^2}$ of small value. For regions with fine texture, images have less spatial correlation and $|C(u,v)|^2$ would obtain a small value at high frequency component i.e. $\sqrt{u^2+v^2}$ of large value. Regions with expression features have relatively rough textures. Important feature information concentrates in low and medium frequency, i.e. abandoning high-frequency component but reserving low and medium frequency in DCT coefficient could still retain most of the information relating to expression features, and then it could screen low and medium frequency features to form feature vector by choosing a larger DCT coefficient.

In order to ensure that low frequency component at top left corner with larger energy could appear before the high-frequency component, the paper adopts triangle extraction[8] (Zig-Zag) whose order of extraction coefficient is as shown in Figure 5.

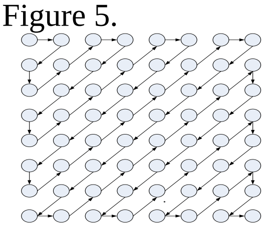


Figure 5 Order of extraction coefficient by zig-zag method

Zig-zag method helps to acquire a larger DCT coefficient to form feature vector which shall be represented as follows.

$$y=[x_0^1, x_1^1, \cdots, x_{K-1}^1, x_0^2, x_1^2, \cdots, \\ x_{K-1}^2, \cdots, x_0^{58}, x_1^{58}, \cdots, x_{K-1}^{58}]^T \tag{2}$$

In formula (2), K represents the number of feature coefficient chosen by Zig-Zag in a feature point image and x_n^m represents the nth feature coefficient in the mth feature point image ($m=1,2,3\cdots 58$; $n=0,1,2,3\cdots K-1$). Studies show that original images could be relatively completely restored by extracting the first 9 feature coefficients in each feature point image, i.e. dimensionality is reduced from the original 24 to 9 through Zig-Zag selection, ensuring feature integrity together with reducing calculated amount in recognition. Therefore, this paper adopts $K=9$ in feature extraction, i.e. the final expression feature vector could be represented as follows.

$$y=[x_0^1, x_1^1, \cdots, x_8^1, x_0^2, x_1^2, \cdots, \\ x_8^2, \cdots, x_0^{58}, x_1^{58}, \cdots, x_8^{58}]^T \tag{3}$$

In Figure 3 (b), region 0 at top left corner of the matrix describes direct component of average grey degree features in each feature point; region 1 describes low frequency component of external features of rough outline of feature points; region 2 describes intermediate frequency components of more detailed features. Therefore, through retaining DCT coefficient of low and medium frequency, it could reduce dimensionality of features and make the extracted expression feature more accurate.

4. Experimental results and analysis

4.1 Clarity and visibility verification of texture features.

As shown in Figure 6, take texture features of any eight key feature points for comparative analysis. The upper are feature maps whose texture features are directly extracted by DCT, and the lower are feature maps extracted after improvement of DT_CWT + DCT. It can be obviously seen that texture features of the lower pictures are much clearer and more obvious than those of the upper pictures, namely the improved algorithm has greatly reduced the impact of feature points position displacement on features extraction. DT_CWT has approximate translation invariance, so that the extracted features appear no aliasing, which means the given sub band has unique z-transform function, so its response is linear and non-time-varying, and the clearly visible images of texture features prove its good approximate translation invariance.

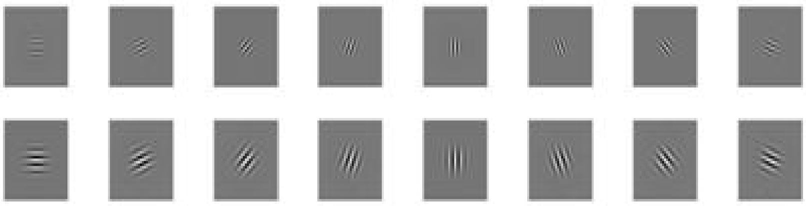
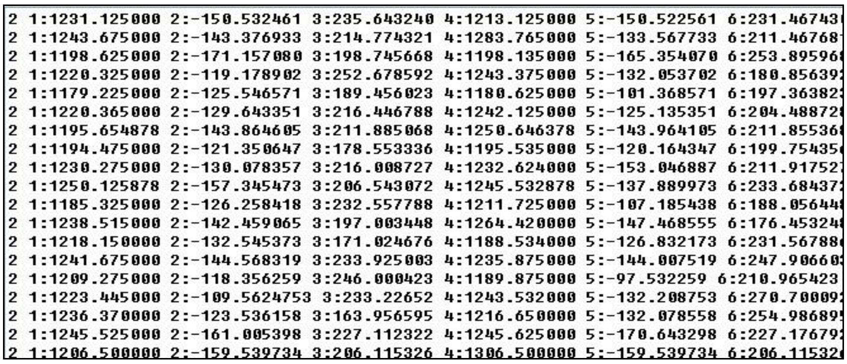


Figure 6 Comparison of texture features before and after improved by DT_CWT

```
2 1:1231.625000 2:-156.728561 3:237.623840 4:1231.625000 5:-156.728561 6:237.623840
2 1:1273.625000 2:-153.427933 3:241.212021 4:1273.625000 5:-153.427933 6:241.212021
2 1:1178.125000 2:-151.678070 3:253.947968 4:1178.125000 5:-151.678070 6:253.947968
2 1:1229.625000 2:-160.080902 3:230.803192 4:1229.625000 5:-160.080902 6:230.803192
2 1:1180.125000 2:-94.960571 3:167.612823 4:1180.125000 5:-94.960571 6:167.612823 7:
2 1:1224.375000 2:-125.139351 3:174.822128 4:1224.375000 5:-125.139351 6:174.822128
2 1:1290.624878 2:-134.920105 3:211.772568 4:1290.624878 5:-134.920105 6:211.772568
2 1:1194.375000 2:-110.163147 3:154.268936 4:1194.375000 5:-110.163147 6:154.268936
2 1:1222.375000 2:-133.048157 3:166.919327 4:1222.375000 5:-133.048157 6:166.919327
2 1:1285.624878 2:-173.881973 3:206.424072 4:1285.624878 5:-173.881973 6:206.424072
2 1:1230.125000 2:-114.134018 3:165.050858 4:1230.125000 5:-114.134018 6:165.050858
2 1:1238.500000 2:-132.878265 3:167.409348 4:1238.500000 5:-132.878265 6:167.409348
2 1:1208.250000 2:-123.543373 3:171.005676 4:1208.250000 5:-123.543373 6:171.005676
2 1:1255.875000 2:-144.755219 3:247.925003 4:1255.875000 5:-144.755219 6:247.925003
2 1:1189.875000 2:-97.750259 3:171.924423 4:1189.875000 5:-97.750259 6:171.924423 7:
2 1:1223.625000 2:-130.365753 3:169.748352 4:1223.625000 5:-130.365753 6:169.748352
2 1:1246.250000 2:-123.046158 3:163.972595 4:1246.250000 5:-123.046158 6:163.972595
2 1:1285.625000 2:-179.045898 3:227.130112 4:1285.625000 5:-179.045898 6:227.130112
2 1:1306.500000 2:-159.539734 3:206.115326 4:1306.500000 5:-159.539734 6:206.115326
2 1:1256.500000 2:-135.220200 3:166.236282 4:1256.500000 5:-135.220200 6:166.236282
2 1:1259.750000 2:-149.367447 3:167.855682 4:1259.750000 5:-149.367447 6:167.855682
2 1:1254.500000 2:-139.652771 3:171.941055 4:1254.500000 5:-139.652771 6:171.941055
2 1:1236.250000 2:-108.827591 3:170.332825 4:1236.250000 5:-108.827591 6:170.332825
2 1:1335.625000 2:-112.860001 3:247.081329 4:1335.625000 5:-112.860001 6:247.081329
2 1:1234.250000 2:-117.101921 3:168.242752 4:1234.250000 5:-117.101921 6:168.242752
```

Figure 7 Partial DCT texture features directly extracted by DTC



```
2 1:1231.125000 2:-150.532461 3:235.643240 4:1213.125000 5:-150.522561 6:231.46743
2 1:1243.675000 2:-143.376933 3:214.774321 4:1283.765000 5:-133.567733 6:211.46768
2 1:1198.625000 2:-171.157080 3:198.745668 4:1198.135000 5:-165.354070 6:253.89596
2 1:1220.325000 2:-119.178902 3:252.678592 4:1243.375000 5:-132.053702 6:180.85639
2 1:1179.225000 2:-125.546571 3:189.456023 4:1180.625000 5:-101.368571 6:197.36382
2 1:1220.365000 2:-129.643351 3:216.446788 4:1242.125000 5:-125.135351 6:204.48872
2 1:1195.654878 2:-143.864605 3:211.885068 4:1250.646378 5:-143.964105 6:211.85536
2 1:1194.475000 2:-121.350647 3:178.553336 4:1195.535000 5:-120.164347 6:199.75435
2 1:1230.275000 2:-130.078357 3:216.008727 4:1232.624000 5:-153.046887 6:211.91752
2 1:1250.125878 2:-157.345473 3:206.543072 4:1245.532878 5:-137.889973 6:233.68437
2 1:1185.325000 2:-126.258418 3:232.557788 4:1211.725000 5:-107.185438 6:188.05644
2 1:1238.515000 2:-142.459065 3:197.003448 4:1264.420000 5:-147.468555 6:176.45324
2 1:1218.150000 2:-132.545373 3:171.024676 4:1188.534000 5:-126.832173 6:231.56788
2 1:1241.675000 2:-144.568319 3:233.925003 4:1235.875000 5:-144.007519 6:247.90660
2 1:1209.275000 2:-118.356259 3:246.000423 4:1189.875000 5:-97.532259 6:210.965423
2 1:1223.445000 2:-109.5624753 3:233.22652 4:1243.532000 5:-132.208753 6:270.70009
2 1:1236.370000 2:-123.536158 3:163.956595 4:1216.650000 5:-132.078558 6:254.98689
2 1:1245.525000 2:-161.005398 3:227.112322 4:1245.625000 5:-170.643298 6:227.17679
2 1:1206.500000 2:-159.539734 3:206.115326 4:1306.500000 5:-159.539734 6:206.11532
```

Figure 8: Partial DCT texture features extracted by DT_CWT+DCT

4.2 Features difference verification.

To further illustrate the impact of position displacement of feature points on features extraction is weakened after improvement of DT_CWT+DCT, the feature difference extracted from the same feature points are compared with partial DCT features extracted before and after improvement of DT_CWT in Figure 7 and 8.

Note: “1” in this figure represents the first row, and the following is DCT feature value, only 6 rows of feature values are listed here, and each row represents the feature value of the same feature point.

As can be seen from Figure 7, the maximum feature value of the first row is 1290.625000, and the minimum is 1178.12500, the difference is more than 112. In the second row, the difference between the maximum and minimum feature value is more than 85, which shows feature value difference of the same feature point extracted by DCT is larger. As can be seen from Figure 8, the maximum feature value of the first row is 1250.125878 and the minimum is 1179.225000, the difference is about 71, which is over 40 less than that before improvement of DT_CWT, indicating that feature value difference of the same displaced feature point has improved after impact of DT_CWT approximate translation invariance, which is more beneficial to improve the recognition rate.

4.3 Recognition rate verification.

In order to demonstrate effectiveness of DT_CWT + DCT algorithm more directly, the recognition rate is verified then. This system firstly collects images through USB camera, and then grays the images and equalizes histogram to reduce the impact of light on facial expression features, and then uses adaboost algorithm for face detection [9], sets the key feature points located by ASM to be the region of interest, extracts the overall range information and improved DT_CWT + DCT texture features of partial expression and conducts weighted fusion, finally, the system classifies the expressions using SVM.

Figure 9 Recognition situations of an experimenter before and after the improvement

Next, recognition rates of facial expressions before and after the improvement are compared. Figure 9 shows the recognition situation of an experimenter before and after the improvement in the same experiment environment. It can be seen that before the improvement, errors occur when recognizing the similar expressions, as shown in the first picture of Figure 9, neutral is mistakenly recognized as happy, which is caused by the inaccurate extraction because of displacement location of feature points by ASM. Expressions in the second row pictures are all correctly recognized, and it can be seen extraction of texture features of partial expressions after improvement has greatly reduced the impact of position displacement on feature points.

Tab.1 Result statistics of facial expression recognition

Expression		surprise	happy	fear	sadness	neutral	disgust	anger
Recognition number		550	550	550	550	550	550	550
Incorrect recognition number	a	38	36	43	52	62	102	125
	b	24	28	37	32	42	71	75
Recognition rate	a	93.1	93.5	92.2	90.5	88.7	81.5	77.3
	b	95.6	94.9	93.3	94.2	92.4	87.1	86.4
Recognition time	a	66.24	60.51	67.39	59.73	62.45	64.96	61.96
	b	67.34	69.56	65.23	62.34	61.45	64.59	68.35
Average recognition rate	a				88.1%			
	b				92.0%			

Note: a: before the improvement; b: after the improvement（Unit of time: ms)

To further compare the recognition rate before and after the improvement, 7 kinds of expressions are experimented for 550 times respectively (the expressions come from JAFFE expressions [10]), the recognition rate statistics are shown in Table 1.

As can be seen from Table 1: when there is not much difference in recognition time, recognition rate has improved by an average of 3.9% after improvement of DT_CWT + DCT, especially for recognition rate of expressions with subtle changes such as anger and disgust, they have improved by 9.1% and 5.6% respectively. The recognition system after improvement also can effectively improve recognition rate for surprise, happiness, fear, sadness, neutrality expressions and so on. Therefore, features extraction after improvement of DT_CWT + DCT has effectively reduced the impact of ASM key features location displacement, and recognition rate of the whole recognition system has been improved to a certain degree, especially for facial expressions with subtle changes.

5. Conclusions

This paper puts forward an ASM extraction method of partial expressions improved by DT_DWT + DCT. A large number of experiments show that this method has greatly improved the problems that partial texture features are unclear and not obvious due to ASM feature point location displacement, and recognition rate of facial expressions has been improved to a certain degree after the improvement.

Reference

[1] Wang Y, De-Qian Y E. Facial Expression Recognition Based on Gabor and Two Times DCT[J]. Microelectronics & Computer, 2009, 26(5):262-264.

[2] Cootes T, Taylor C. Active shape models-smart snakes. BMVC[J]. British Machine Vision Conference, 1992:266-275.

[3] Ying Z L, Jing-Wen L I, Zhang Y W. Facial Expression Recognition Based on SLLE with Expression Weighted Distances[J]. Pattern Recognition & Artificial Intelligence, 2010, 23(2):278-283.

[4] Shan C, Gong S, Mcowan P W. Facial expression recognition based on Local Binary Patterns: A comprehensive study[J]. Image & Vision Computing, 2009, 27(6):803-816.
[5] Zhu S P. Facial expression recognition based on ASM and Multi-instance boosting[J]. Computer Modelling and New Technologies, 2014,18(12):323-330.
[6] Jung S U, Kim D H, An K H, et al. Efficient rectangle feature extraction for real-time facial expression recognition based on AdaBoost[J]. Intelligent Robots & Systems .ieee/rsj International Conference on, 2005:1941-1946.
[7] Kingsbury N G. Complex wavelets for shift invariant analysis and filtering of signals [J].Appl.Comp.Harmonic Anal., 2000, 10(3):234-253.
[8] Chen L, Zhou C, Shen L. Facial Expression Recognition Based on SVM in E-learning[J]. Ieri Procedia, 2012:781-787.
[9] Lozano-Monasor E, López M T, Fernández-Caballero A, et al. Facial Expression Recognition from Webcam Based on Active Shape Models and Support Vector Machines[J]. Lecture Notes in Computer Science, 2014:147-154.
[10] Cheng F, Yu J, Xiong H. Facial Expression Recognition in JAFFE Dataset Based on Gaussian Process Classification[J]. IEEE Transactions on Neural Networks, 2010, 21(10):1685-1690.