行政院國家科學委員會專題研究計畫 成果報告

以視覺為基礎之人類行為分析系統
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2221-E-216-045-
執 行 期 間 ：96 年 08 月 01 日至 97 年 07 月 31 日
執 行 單 位 ：中華大學資訊工程學系

計 畫 主 持 人：連振昌
共 同 主 持 人：石昭玲
計畫參與人員：碩士班研究生-兼任助理人員：蔡明修
碩士班研究生-兼任助理人員：石正崙

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，1 年後可公開查詢

中 華 民 國 97 年 10 月 09 日

行政院國家科學委員會補助專題研究計畫 ☑ 成 果 報 告 □期中進度報告

（計畫名稱）

以視覺為基礎之人類行為分析系統

計畫類別：☑ 個別型計畫　　□ 整合型計畫
計畫編號：NSC　96-2221-E-216 -045-
執行期間：　96　年　8　月　1　日至　97　　年　7　月　31　日

計畫主持人：連振昌
共同主持人：石昭玲
計畫參與人員：　蔡明修、石正崙

成果報告類型(依經費核定清單規定繳交)：☑精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
☑出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，☑一年□二年後可公開查詢

執行單位：中華大學資訊工學系

中　華　民　國　97　年　10　月　6　日

一、中文摘要

近年來視訊監控隨著電腦視覺技術之精進以及對居家照護、犯罪預防的強烈需求，因此發展一套以電腦視覺為基礎之目標物特寫追蹤系統，同時對所監控之目標物做各式的行為分析之系統有其重要性及迫切性。本三年期研究計畫目標為建置一套以電腦視覺為基礎之目標物特寫追蹤系統同時對所監控之目標物做各式的行為分析。於本計劃所規劃之行為分析包括二大部分:肢體行為分析及臉部行為分析。肢體行為分析主要研究人的步伐及動作分析。臉部行為分析著重於辨識臉部及眼睛表現出之表情。本計劃分成三年進行，各年度執行項目分別為雙相機之多目標物特寫追蹤系統、全方位之人類步伐分析及於低解析度影像上之臉部表情分析系統。於第三年計畫執行時結合 95 年度之國科計畫(NSC-95-2211-E-216-030-)中所研究之動作偵測分析技術，同時整合肢體行為及臉部行為來分析所觀測之人發生什麼事件，例如在居家照護環境下發現被照護人跌倒，隨即偵測被照護者表情判斷此事件是否須發出警訊。或者於大型的監控環境中(例如銀行大廳)監控是否有人入侵禁制區(櫃檯)，隨即記錄臉部影像及步伐特徵，作為日後偵防的輔助工具。計劃於此三年期計畫發展出一套以視覺為基礎之人類行為分析系統。**此計劃原提三年期計畫，因只核定為單年期計畫，所以只執行於低解析度影像上之臉部表情分析系統。**

關鍵字：目標物特寫追蹤、肢體行為分析、步伐及動作分析、臉部行為分析

二、英文摘要

In this study, we propose a novel image-based facial expression recognition method called “expression transition” to identify six kinds of facial expressions (anger, fear, happiness, neutral, sadness, and surprise) at low-resolution images. The boosted tree classifiers and template matching are used to locate and crop the effective face region that may characterize the facial expressions. Then, the expression transformed images via a set of expression transition matrices are matched with the real facial images to identify the facial expressions. The proposed system can recognize the facial expressions with the speed of 0.24 seconds per frame and accuracy above 86%.

Keywords: facial expression recognition, expression transition matrices

三、報告內容

[1] **Cheng-Chang Lien,** Lin-Yi Lin and Cho-Ming Tu, “A New Appearance-Based Facial Expression Recognition System with Expression Transition Matrices,” IEEE 3rd International Conference on Innovative Computing, Information and Control (ICICIC 2008), June 18-20, 2007, Dalian, China. **(EI)**

(此篇論獲大會推薦刊登於 IJICIC)

On behalf of ICICIC2008, it is our great pleasure to inform you that
>> your paper (A New Appearance-Based Facial Expression Recognition System
>> with Expression Transition Matrices//) has been recommended to be
>> published in the International Journal of Innovative Computing,
>> Information and Control (IJICIC) (indexed by SCI, http://www.ijicic.org)
>> for the Special Issue, "*New Trends in Information Processing and
>> Applications*".

A New Appearance-Based Facial Expression Recognition System with Expression Transition Matrices

Cheng-Chang Lien, Lin-Yi Lin and Cho-Ming Tu
Department of Computer Science and Information Engineering,
Chung Hua University, Hsinchu, Taiwan
E-mail: cclien@chu.edu.tw

Abstract

In this study, we propose a novel image-based facial expression recognition method called "expression transition" to identify six kinds of facial expressions (anger, fear, happiness, neutral, sadness, and surprise) at low-resolution images. The boosted tree classifiers and template matching are used to locate and crop the effective face region that may characterize the facial expressions. Then, the expression transformed images via a set of expression transition matrices are matched with the real facial images to identify the facial expressions. The proposed system can recognize the facial expressions with the speed of 0.24 seconds per frame and accuracy above 86%.

1. Introduction

Generally, facial expression recognition systems are roughly categorized into feature-based [1], image-based [2], and model-based [3] methods. In the feature-based methods, the shapes and locations of eyebrows, eyes, nose, and mouth are extracted to form the expression feature vectors. Trujillo et al. [1] propose an interest point detection method to localize the facial features using intensity variations of OTCBVS IRIS thermal image [8]. In the image-based methods, the holistic spatial analysis and local spatial analysis are used to recognize the facial expression. Donato et al. [2] explore many new approaches for the facial image representation to recognize action units (AUs) [9], which include principal component analysis (PCA), independent component analysis (ICA), local feature analysis (LFA), and linear discriminant analysis (LDA). In the model-based methods, the statistical model is constructed from training images and used to recognize the facial expressions. Huang et al. [3] apply the point distribution model and gray-level model to find the facial expression features.

To recognize the facial expressions accurately, several pattern recognition methods are integrated in the facial expression recognition systems. In [4], machine learning methods including adaboost, linear discriminant analysis (LDA), and support vector machines (SVM) are applied to recognize the facial expressions and the accuracy of these methods are analyzed. However, the above-mentioned researches have two significant drawbacks. Firstly, the feature extraction methods can not extract the facial features (shape, color, and position) robustly because of hair occlusion, lighting variation, and wrinkle. Secondly, most of current researches [1-4] are addressed on the facial expression recognition at high-resolution images. In this paper, a novel method called "expression transition" is developed to recognize the facial expression at low-resolution images. In addition, the illumination compensation projection method [11], template matching and Hough transform [12] are used to improve the locating of eyes and mouth. Based on the positions of eyes and mouth the effective facial region can be extracted for the expression transformation process. The system block diagram is shown in Fig. 1.

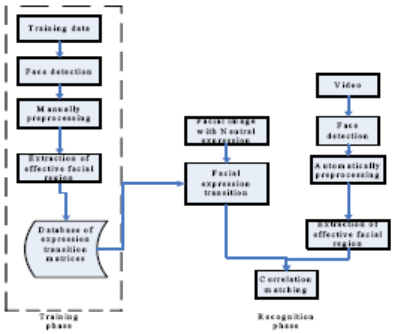

Fig. 1. The proposed facial expression recognition system

2. Image preprocessing

Before recognizing the facial expressions, some image preprocesses show in Fig. 2 are needed to extract an effective facial region.

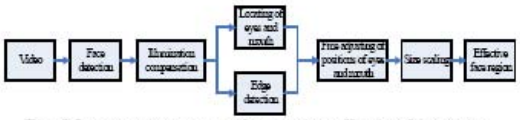

Fig. 2 Image preprocesses of extracting effective face region.

2.1 Face detection and illumination compensation

The face region is detected by a cascade of boosted tree classifiers [13]. To overcome the influence of illumination variation, the illumination compensation [11] is applied. By using the multiple regression model defined in (1), we can find the illumination compensation plane.

$$Y = XB + \varepsilon \quad (1)$$

where $\boldsymbol{X}$ is the image coordinate matrix, $\boldsymbol{Y}$ is the image intensity, $\boldsymbol{B}$ is the regression parameters and $\boldsymbol{\varepsilon}$ is the fitting error. The illumination compensation is illustrated in Fig. 3.

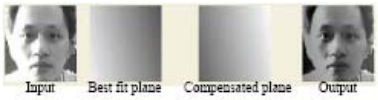

Fig. 3. Multiple regression model for illumination compensation.

2.2 Locating of eyes and mouth

Position of eye and mouth are locating by using the template matching and edge information. Because the pupil is especially conspicuous in red channel, a 5×5 template is used to find the position of pupil. However, the pupil position obtained from the template matching can only serve as the candidate position of the pupil. Hence, the Hough transform is applied to find the center position of pupil from the edge information extracted from the canny edge. Based on the candidate position, different radiuses are used to find the precise position of pupil. To find the position of the mouth, the color template matching and edge searching are applied again. The algorithm of locating the mouth is described in the following steps.

1. Based on the anthropometry and center of pupils, a line searching segment shown in Fig. 4 is determined.
2. Compute the edge density along the searching line segment. If the edge density is larger than a predefined threshold, then this point is categorized into the candidate mouth positions.
3. Apply the template matching to find the precise position of the mouth.

Fig. 4(b) illustrates the extracted positions of eyes and mouth.

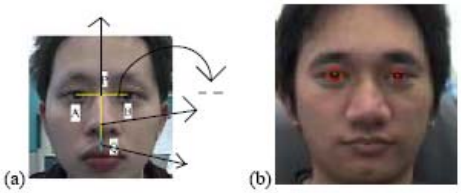

Fig. 4 (a) The blue line segment used to search the position of mouth. (b) The extracted positions of eyes and mouth.

Finally, the effective facial region is extracted according to the coordinates of eyes and mouth. Furthermore, the inclination correction and size scaling are performed to extract the effective face region shown in Fig. 5 for recognizing the facial expressions.

Fig. 5. Effective face regions for recognizing the facial expressions.

3. Expression Transition Matrix

Here, we propose an image-based facial expression recognition method called "expression transition" to identify six kinds of facial expressions with high efficiency and accuracy at low-resolution images.

3.1. Direct mapping

For each normalized image, an expression feature vector **a** is constructed by scanning the image from left-top to right-bottom and written as

$$\mathbf{a} = \left[a_{0,0} \quad a_{0,1} \quad \cdots \quad a_{n,n}\right]^T, \quad (2)$$

where $a_{x,y}$ denotes the pixel value at position (x, y). We employ direct mapping approach [10] to calculate the expression transition matrices shown in Fig. 6.

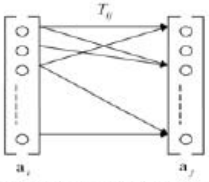

Fig. 6. Facial expression transition using direct mapping.

Given two subsets of feature vectors with facial expression indices i and j, the transformation between two facial expressions may be modeled as

$$B_j = T_{ij} B_i, \quad (3)$$

where $B_j = [a_j^1, a_j^2 \cdots a_j^K]$ and $B_i = [a_i^1, a_i^2 \cdots a_i^K]$, k is the number of expression feature vectors formed by individual persons and T_{ij} represents the expression transition matrix from facial expression indices i to j.

3.2. Expression transformation

Once the expression transition matrices are derived, the facial expression transformation from expression i to j can be written as:

$$a_j = T_{ij} a_i . \quad (4)$$

Based on the formula in (4), we may simulate some expression transformed facial images shown in Fig. 7.

Fig. 7. Expression transformed facial images for (a) the images in Cohn-Kanade database and (b) the captured facial images.

4. Facial Expression Recognition

Based on the set of expression transition matrices $T = \{T_{neutral\text{-}anger}, T_{neutral\text{-}week\ anger}, T_{neutral\text{-}fear}, T_{neutral\text{-}week\ fear}, T_{neutral\text{-}happiness}, T_{neutral\text{-}week\ happiness}, T_{neutral\text{-}neutral}, T_{neutral\text{-}sadness}, T_{neutral\text{-}week\ sadness}, T_{neutral\text{-}surprise}$ and $T_{neutral\text{-}week\ surprise}\}$, we can transform the facial image from the Neutral facial expression to the desired facial expression. Then, the method of correlation matching defined in Eq. (5) is applied to calculate the correlation coefficients between transformed facial images and the real facial images and then the facial expression can be recognized.

$$\gamma = \frac{\sum_x \sum_y [\hat{f}(x,y) - \hat{f}_m][f(x,y) - f_m]}{\left\{\sum_x \sum_y [\hat{f}(x,y) - \hat{f}_m]^2 \sum_x \sum_y [f(x,y) - f_m]^2\right\}^{\frac{1}{2}}} \quad (5)$$

where $\hat{f}$ is transformed facial image, f is the registered facial image with Neutral facial expression, and $\hat{f}_m$ and f_m are average pixel values for $\hat{f}$ and f respectively.

In general, each facial expression can have different degrees of facial actions, e.g., happiness may be a hearty laugh or smile. Hence, two transition matrices for the expression transition from neutral to happiness, anger, fear, sadness, or surprise are generated to simulate the expression transition with different degrees. Then, each facial expression is recognized by fusing two correlations obtained by matching the input facial image with the simulated facial image of different degrees. The fusion rule is defined as:

$$R = \sqrt{r * r'} . \quad (6)$$

Fig. 8 shows the block diagram of the proposed facial expression recognition system.

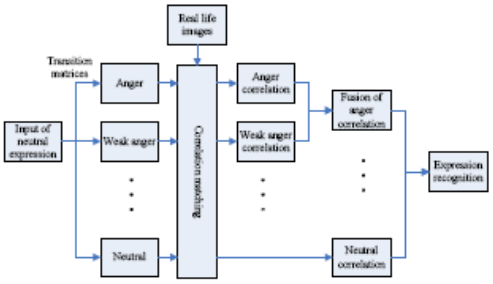

Fig. 8. The block diagram of the proposed facial expression recognition system.

5. Experimental Results

The proposed facial expression recognition system consists of two phases: training phase and recognition phase. In the training phase, 240 facial images captured from 104 persons in Cohn-Kanade facial expression database [14] are used to construct the expression transition matrices. The effective face region are extracted by the method mentioned in section 2 and normalized to a low-resolution image with size 30×30. Then, the direct mapping is applied to construct the expression transition matrices. In the recognition phase, 20 video clips from 8 different persons are used as the test data to evaluate the proposed facial expression recognition system. Each facial expression can be recognized with 0.24 seconds per frame. Table 1 shows the accuracy analysis with temporal error filtering of facial expression recognition system. The average recognition accuracy is above 86%. Because the facial expressions (Happiness) in video clips 8 and 18 are not obvious shown in Fig. 9, they are misclassified into Neutral expression. In addition, some images with eye

closing shown in Fig. 10 also make the expression recognition inaccurate.

Table 1. The accuracy analysis of the proposed facial expression recognition systems.

Human ID	video clip number	neutral right	neutral wrong	correct rate	happy right	happy wrong	correct rate	anger right	anger wrong	correct rate	surprise right	surprise wrong	correct rate
1	video 1	32	4	0.89	15	1	0.94	10	1	0.91	13	2	0.87
1	video 2	58	4	0.94	25	4	0.86	20	2	0.91	16	2	0.89
1	video 3	75	3	0.96	20	3	0.87	21	4	0.84	15	4	0.79
1	video 4	69	5	0.93	14	2	0.88	31	2	0.94	19	2	0.90
2	video 5	26	1	0.96	22	3	0.88	20	3	0.87	15	1	0.94
2	video 6	25	9	0.74	18	2	0.90	20	5	0.80	25	2	0.93
2	video 7	36	3	0.92	23	1	0.96	17	2	0.89	17	2	0.89
3	video 8	30	0	1.00	14	12	0.54	17	2	0.89	25	0	1.00
3	video 9	28	4	0.88	23	1	0.96	18	3	0.86	24	1	0.96
3	video 10	33	1	0.97	21	2	0.91	30	4	0.88	18	2	0.90
4	video 11	33	0	1.00	25	2	0.93	22	2	0.92	26	1	0.96
4	video 12	28	2	0.93	15	3	0.83	24	3	0.89	23	2	0.92
5	video 13	39	1	0.98	21	2	0.91	18	2	0.90	22	2	0.92
5	video 14	13	0	1.00	18	2	0.90	25	1	0.96	31	1	0.97
5	video 15	55	3	0.95	10	2	0.83	11	1	0.92	16	1	0.94
6	video 16	24	14	0.63	30	0	1.00	21	1	0.95	25	1	0.96
6	video 17	30	8	0.79	34	3	0.92	18	1	0.95	25	2	0.93
7	video 18	69	1	0.99	14	11	0.56	30	5	0.86	20	3	0.87
7	video 19	36	2	0.95	17	5	0.77	14	4	0.78	13	4	0.76
8	video 20	56	2	0.97	24	4	0.86	15	1	0.94	43	11	0.80
	total	795	67	0.92	403	65	0.86	402	49	0.89	431	46	0.90

Fig 9. Facial expressions (Happiness) in video clips 9 and 18 are not obvious. They are misclassified as Neutral expression.

Fig. 10. Eye closing makes the expression recognition inaccurate.

6. Conclusion

In this paper, we propose a novel image-based facial expression recognition method called "expression transition" to identify six kinds of facial expressions including anger, fear, happiness, neutral, sadness, and surprise. The experimental results show that the proposed facial expression recognition system can recognize the facial expressions with the speed of 0.24 seconds per frame and accuracy above 86%.

Acknowledgement

This work has been partially funded by National Science Council under project no. NSC 96-2221-E-216-45.

7. References

[1] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "Automatic feature localization in thermal images for facial expression recognition", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 3, pp. 14.

[2] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, vol. 21, pp. 974-989.

[3] C. L. Huang and Y. M. Huang, "Facial expression recognition using model-based feature extraction and action parameters classification", Journal of Visual Communication and Image Representation, 1997, vol. 8, pp. 278-290.

[4] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior", IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2005, vol. 2, pp. 568-573.

[5] R. Cowie, E. Dougls-Cowie, N. Tsapatsoulis, G.Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction", IEEE Signal Processing Magazine, 2001, vol. 18, pp. 32-80.

[6] P. Ekman, Telling lies: clues to deceit in the marketplace, politics, and marriage, W. W. Norton, New York, 1985.

[7] M. Heller and V. Haynal, "The faces of suicidal depression (translation les visages de la depression de suicide)", Kahiers Psychiatriques Genevois (Medicine et Hygiene Editors), 1994, vol. 16, pp. 107-117.

[8] OTCBVS database, ttp://www.cse.ohiostate.edu/otcbvs - bench.

[9] P. Ekman and W. Friesen, Facial action coding system: a technique for the measurement of facial movement, Consulting Psychologists Press, Palo Alto, 1978.

[10] A. Utsumi and N. Tetsutani, "Adaptation of appearance model for human tracking using geometrical pixel value distributions", Proceedings of Sixth Asian Conference on Computer Vision, 2004, vol. 2, pp. 794-799.

[11] J. Ko, E. Kim, and H. Byun, "A simple illumination normalization algorithm for face recognition", Springer-Verlag Berlin Heidelberg 2002, PRICAI 2002, LNAI 2417, pp. 532-541, 2002.

[12] Z. Xu and P. Shi, "A robust and accurate method for pupil features extra," 18th International Conference on Pattern Recognition, vol. 1, 20-24 Aug. 2006. pp. 437–440.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2001, vol. 1, pp. 511-518.

[14] Cohn-Kanade facial expression database, http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

95 年 10 月 16 日

附件三

報告人姓名	連振昌	服務機構及職稱	中華大學資訊工程學系 副教授
時間 會議 地點	June 18-20, 2007, Dalian, China	本會核定 補助文號	NSC 96-2221-E-216-045-
會議名稱	（中文）IEEE 第三屆國際創新計算、訊息與控制會議 （英文）IEEE 3rd International Conference on Innovative Computing, Information and Control (ICICIC 2008)		
發表論文題目	（中文）一個使用表情轉移矩陣之表情辨識系統 （英文）A New Appearance-Based Facial Expression Recognition System with Expression Transition Matrices		

一、參加會議經過

本人於2007年6月18-20日赴大陸大連市參加 “IEEE 3rd International Conference on Innovative Computing, Information and Control (ICICIC 2008)” 國際會議，會中發表論文一篇，如下所示：

Cheng-Chang Lien, Lin-Yi Lin and Cho-Ming Tu, “A New Appearance-Based Facial Expression Recognition System with Expression Transition Matrices,” IEEE 3rd International Conference on Innovative Computing, Information and Control (ICICIC 2008), June 18-20, 2007, Dalian, China. **(EI)**

(此篇論獲大會推薦刊登於 IJICIC)

On behalf of ICICIC2008, it is our great pleasure to inform you that
>> your paper (A New Appearance-Based Facial Expression Recognition System
>> with Expression Transition Matrices//) has been recommended to be
>> published in the International Journal of Innovative Computing,
>> Information and Control (IJICIC) (indexed by SCI, http://www.ijicic.org)
>> for the Special Issue, "*New Trends in Information Processing and
>> Applications*".

二、與會心得

會議中與各國學者作深切的學術交流，獲益良多。

三、考察參觀活動(無是項活動者省略)

無

四、建議

無

五、攜回資料名稱及內容

ICICIC2008 論文集