行政院國家科學委員會補助專題研究計畫第一年成果報告

總計畫：結合生物反饋之新世代腦機介面及其在移動載具
　　　　控制之應用
子計畫四：以機器視覺為基礎之人意向移動載具獲知與追
　　　　　蹤

　　　計畫主持人：張志永
　　　共同主持人：

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

中　華　民　國　97　年　9　月　30　日

行政院國家科學委員會專題研究計畫報告

總計畫：結合生物反饋之新世代腦機介面及其在移動載具控制之應用

子計畫四:以機器視覺為基礎之人意向移動載具獲知與追蹤

Subproject IV: A Machine Vision-Based Human-Intending Moving Vehicle Capture and Tracking

主持人：張志永　　交通大學電機與控制工程學系教授

一、 中文摘要

　　本計畫以機器視覺為基礎之人意向移動載具獲知與追蹤課題。腦波生理訊號，通常受試驗者的視覺、聽覺、感受等影響，本子計畫的目的，在於利用視覺狀態回饋給使用者，探討其對的人腦波訊號之影響。利用本計畫結果，子計畫三可據此鑑別出重要視覺腦波生理訊號的位置，即頻道，及訊號強弱變化，一起與子計畫五方向感腦波訊號、子計畫二注意力腦波訊號，整體訓練學習後，能提昇在移動載具的控制效率。本子計畫執行所需之移動載具，擬以電動代步車為實驗平台。第一年將探索在室內環境情況下，人坐在移動載具車之視訊獲知及其方位獲知方法。

　　由於近年來智慧型機器人的發展迅速，機器人與人之間的互動也愈來愈頻繁，當機器人與人溝通時，必須知道人的位置、距離及面對的方向等資訊。在此篇論文中，我們結合二維影像輪廓比對與臉部方位偵測來完成三維人體方位偵測。首先，任一張影像的前景人物利用一個基於前後影像比值而建立之統計背景模型抽取出來，並將抽取出來影像轉換成二值化的影像格式，進而獲得前景人物的輪廓。透過人體輪廓樣板比對及線性內插法，可以初步得到前景人物所朝的方向。當人臉朝向前方 30° 以內時，也可透過雙眼與臉的三角幾何關係來估算人體所朝的方向。經實驗證明，我們提出的方法對於人體方位偵測的準確度相當高，角度誤差低於 4° 。

關鍵詞：人物方位偵測、輪廓對應、臉部與眼睛抽取、傅利葉描述器

英文摘要

This project is concerned with machine vision-based moving vehicle capture and tracking. It is known that brain signals are dependent on the excitations of vision, audio, and feeling of a tester. The purpose of this project is to provide various feedback to a user and then investigate its effect on brain signaling. With an excitation, subproject III will identify the responsive brain channel and its signaling variation. The brain signaling variations of a person invoking a multi-task distraction and spatial disorder are studied by subprojects II and V, respectively. With these identified responsive brain

signaling, multi-mode bio-feedback embedded in brain-computer interface formulation can reach a new era.

In recent years, the advancement of robot technology brings robot into human daily activities, and makes robot as an essential part of modern life. The interaction between human and robot has become more and more frequent. In this year test-bed, the man will be sitting in a car to interact with a robot; i.e., his orientation will be recognized by the robot. When robot interact with the human, the information of the position, distance, and direction of a human becomes important. we combine 2D image silhouette matching and face direction detection to detect a human direction. When one's face is within 30°, one's pupils and their geometric relationship can also be exploited to estimate one's direction. By numerical simulation, we have obtained a high accuracy, less than 4° on the average, in subject direction detection.

Keywords: Subject Direction Detection, Silhouette Matching, Face and Eye Extraction, Fourier Descriptor.

I. Introduction

The human visual system has an uncanny ability to recognize objects from single views, even when presented monocularly under a fixed viewing condition. However, for computer vision, it is a very challenging task to recognize objects in different directions from 2D image. Therefore, model construction and shape recognition of 3-D objects from 2-D silhouettes have been important areas in computer vision.

In the paper, we present a technique to detect a 3-D subject direction from silhouettes of subject from multiple viewpoints, and combine the direction of subject's face, when the face being within $\pm 30^\circ$ forward, to estimate subject direction. We recovered the subject orientation from the silhouette of frontal view, which was taken from the PTZ camera. Then we use 2-D contour matching process and linear interpolation to estimate the direction of subject. In order to enhance the accuracy of direction detection, we also further fuse the face direction detection if face is available and suitably large. As a consequence, we can identify the known subject's orientation.

II. Subject Direction Detection

We propose an approach which combines Fourier Descriptor (FD) based on Fourier series analysis, with Linear Discriminate Analysis (LDA). This method can be used to extract subject features and to optimize the class separability of different subjects by their contours. First, we utilize Fourier descriptors to discriminate different direction of target subject. Secondly, we use LDA to maximize between-class and in the same time minimize within-class variations to improve the classification performance. At last, we calculate the Euclidean distance between the subject of input image and our model which built from Fourier descriptors to estimate the direction angle of the subject.

A. Subject Extraction

We propose to utilize the frame ratio to build the background model. Each pixel of background scene is characterized by three

statistics: minimum intensity value $n(x, y)$, maximum intensity value $m(x, y)$ and maximum inter-frame ratio $d(x, y)$ of a background video. Because these three values are statistical and obtained from a background modeling algorithm [1], we need a background video, without any moving objects, for background model training. Let I be an image frame sequence and contains N consecutive images. $I_i(x, y)$ be the intensity of a pixel which is located at (x, y) in the i-th frame of I. The background model, $[m(x, y), n(x, y), d(x, y)]$, of a pixel is obtained by

$$\begin{bmatrix} m(x,y) \\ n(x,y) \\ d(x,y) \end{bmatrix} = \begin{cases} \begin{bmatrix} \max_i \{I_i(x,y)\} \\ \min_i \{I_i(x,y)\} \\ \max_i \{I_i(x,y)/I_{i-1}(x,y)\} \end{bmatrix} & \text{if } I_i(x,y)/I_{i-1}(x,y) \ge 1 \\ \begin{bmatrix} \max_i \{I_i(x,y)\} \\ \min_i \{I_i(x,y)\} \\ \max_i \{I_{i-1}(x,y)/I_i(x,y)\} \end{bmatrix} & \text{otherwise} \\ & i = 1, 2, \ldots, N. \end{cases}$$

(1)

Foreground subjects can be segmented from every frame of the video stream. Each pixel of the video frame is classified to either a background or a foreground pixel by the difference between the background model and a captured image frame. We utilize the maximum intensity $m(x, y)$, minimum intensity $n(x, y)$ and maximum inter-frame ratio $d(x, y)$ of the training background model to segment a foreground by

$$B(x,y) = \begin{cases} 0, \text{ a background pixel} & \text{if} \begin{cases} I_i(x,y)/m(x,y) < kd(x,y) \\ \text{or} \\ I_i(x,y)/n(x,y) < kd(x,y) \end{cases} \\ 255, \text{ a foreground pixel} & \text{otherwise.} \end{cases}$$

(2)

where $I_i(x, y)$ be the intensity of a pixel which is located at (x, y), $B(x, y)$ is the gray level of a pixel in a binary image and k is a threshold. Threshold k is determined by experiments according to the environments. The value of k affects the mount of information retained in binary image B.

B. Silhouette Matching Using Fourier Descriptor and Linear Discriminant Analysis

Fourier descriptor [2] is a useful implement for describe closed curve shape which obtained from subject contour. We can obtain a periodic function along the closed curve and this function can represent by a Fourier series. In this paper, we introduce Zahn and Roskies' [3] Cumulative-Angle Approach to represent Fourier descriptors. We assume γ is a clockwise-oriented simple closed curve with parametric representation $(x(l), y(l))$. Let $(x(0), y(0))$ be the starting point and we denote $Z(l) = x(l) + jy(l)$, where l is the arc length of the starting point to $Z(l)$ and $0 \le l \le L$. Denote the angular direction of γ at point l by the function $\theta(l)$ and let $\delta_0 = \theta(0)$ be the absolute angular direction at the starting point $Z(0)$. We now define the cumulative angular function $\phi(l)$ as the net amount of angular bend between starting point and point l. With this definition $\phi(0) = 0$ and $\phi(l) + \delta_0$, is identical to $\phi(l)$ except for a possible multiple of 2π. Besides, It is not hard to see that $\phi(L) = -2\pi$ because all smooth simple closed curves with clockwise orientation have a net angular bend of -2π. As a result, $\phi(L)$ does not convey any shape

information. The domain of definition $[0, L]$ of $\theta(l)$ simply contains absolute size information and we would like to normalize to the interval $[0, 2\pi]$ which is standard for periodic functions. Hence we define a normalized variant $\phi^*(t)$ whose domain is $[0, 2\pi]$ and such that $\phi^*(0) = \phi^*(2\pi) = 0$, and expand $\phi^*(t)$ is a periodic function as a Fourier series

$$\phi^*(t) = \mu_0 + \sum_{k=1}^{\infty}(a_k \cos kt + b_k \cos kt) = \mu_0 + \sum_{k=1}^{\infty} A_k \cos(kt - \alpha_k),$$

(3)

Coefficient pair (A_n, α_n) are the polar coordinates for the point (a_n, b_n) can obtain as follows.

$$\mu_0 = -\pi + \frac{1}{L}\sum_{k=1}^{m} l_k \Delta\phi_k,$$ (4)

$$a_n = \frac{-1}{n\pi}\sum_{k=1}^{m}\Delta\phi_k \sin\frac{2\pi n l_k}{L},$$ (5)

$$b_n = \frac{1}{n\pi}\sum_{k=1}^{m}\Delta\phi_k \cos\frac{2\pi n l_k}{L},$$ (6)

where

$$l_k = \sum_{i=1}^{k}\Delta l_i.$$

It is clear from these expressions alone that the Fourier coefficients (a_n, b_n) contain no information relating to absolute position or rotational orientation of the curve. Coefficient pair (A_n, α_n) are the polar coordinates for the point (a_n, b_n). Coefficient A_n is called the n-th harmonic amplitude and α_n is the nth harmonic phase angle.

In addition, we utilize the Linear Discriminant Analysis (LDA) [4] to optimize the class separability and improve the classification performance. LDA is a classic method of classification which seeks to find a linear transformation $\mathbf{w}$ by maximizing the between-class variance and minimizing the within-class variance, has proved to be a suitable technique for discriminating different pattern classes. After finding $\mathbf{w}$, we can map the input data which from FD through $\mathbf{w}$ to a new coordinate $\mathbf{z}$ that has a larger separability. By using this basis, each point in FD can be further projected to another point in this new space by

$$\mathbf{z}_{i,j} = \mathbf{W}y_{i,j}$$ (7)

C. Subject Direction Estimation by Interpolation

After the FD and LDA processes on the input image. We can calculate the Euclidean distance between the subject of interests and built model of the subject direction as shown in Fig. 1. Then we calculate the Euclidean distances between the subject of interests and the seven direction models. The angels of the least two Euclidean distance models could represent the possible direction range of the subject. The direction of minimum distance model is more probable than the direction of the second minimum model, in which the leverage is depending on the ratio of these two minima. The subject's direction could be estimated by the linear interpolation on these two direction angles.
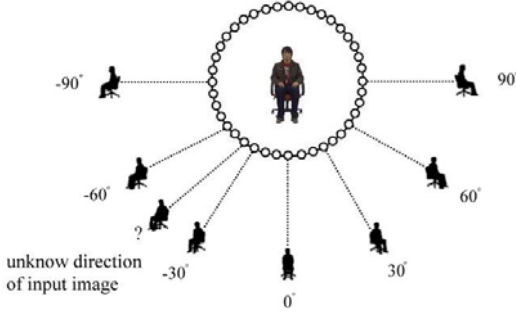
Fig. 1.   The representation of a 3D subject by 2D views.

.

III. Face Direction Detection

In our research, we propose a method to estimate the face direction of subject interests to enhance the accuracy of direction detection of target subject interests. Using image processing, the face angle will be estimated from the geometric relationship of pupils in the face. Firstly, we utilize skin detection formula in the YC_bC_r color space and edge detection to find head region. Secondly, we utilize PCA to find eyes position and then calculate the pupil centers for estimate the face direction. Consequently, we can combine object direction from FD and face direction to better detect the direction of subject/object of interests.

A. Face Tracking by PTZ Camera

In our research, we using Pan-Tilt-Zoom (PTZ) camera to track human face before face direction detection. In order to estimate face direction correctly, we have to zoom the face of target person in to occupy the image between 70% to 80% vertically. We utilize YC_bC_r skin color segmentation method to detect the face region initially. There will be some noise or other skin color region (e.g. hand, leg, etc.)

segmented left in the extracted image. We have to eliminate noise and non-face region. Then calculate the coordinates of the center of the face region, which are used to compute the distance to the center of image. We recode the pixels of x-direction and y-direction from the distance which is proportionality to the camera pan and tilt time, and compute the occupying ratio of the human face in the image. According to these parameters, PTZ camera will track human face automatically.

B. Head Region Extraction

The first stage of the algorithm is to classify the pixels of the input image into skin region and non-skin region [5-9]. To do this, we obtain a skin-color reference map in YC_bC_r color space [5]. The second stage is to combine the region of skin color detected and the result from edge detected. Because skin color segmentation is unable to get the whole region of head, thus we have to utilize the edge detection method, then do the union with the face region. Thus, we can get a more complete head region.

C. Eye Detection   and Pupil Center Estimation Using PCA

Most approaches in computer recognition of faces and expressions have been focused on detecting individual features such as eyes, head outline, mouth, or defining a face model by position, size, and relationships among these features [10-12]. Features extraction plays an essential role in the pre-processing stage. Principal component analysis (PCA) has been commonly used to face recognition problems. Typical PCA algorithm is one of the main streams of research on face feature processing

[6]. PCA has advantage over other face recognition schemes in its speed and simplicity. We utilized PCA in the pre-processing stage to extract features from input eye region image which has been extracted in last Section.

The next step is to find the pupil center of the eye region found by PCA. We use intensity to segment pupil region as shown in Fig. 2(b). Observe the region of iris in Fig. 2(b), the segmentation through pixel intensity thresholding have roughly found out the area of the pupil. But there exist a few low intensity pixels having segmented between iris and sclera, we utilize a rectangular mask to search the iris position. Moving the mask from top-left to down-right to record the number of pixel value is 1 in the mask. The position of the mask with largest number of 1 in the rectangular would be the best iris location as shown in Fig. 2(c). Therefore, the center of the mask is nearly and hence identified to be the center of the iris.
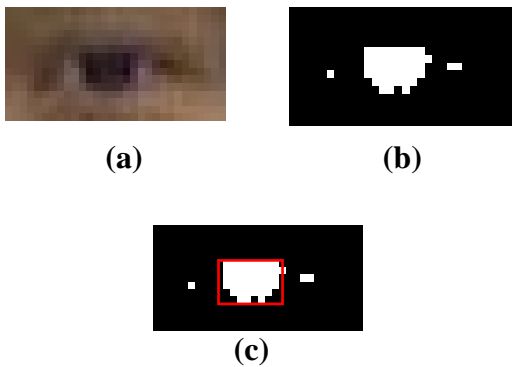

Fig. 2. (a) Input image. (b) Intensity field. (c) The mask with the largest number of pupil pixels.

D. Face Direction Estimation

After we calculate out the pupil center, face direction angle can estimated by using face and head model as shown in Fig. 3. Since the nose position is located near by the center of the eyes. We select the center of two pupils as the nose position on the horizontal axis. Therefore, the face direction angle ψ can be obtained by the following equation.

$$\begin{cases} a = (R + L) / 2 \\ b = (R + L) / 2 - R \end{cases} \Rightarrow \sin \psi = \frac{b}{a} = \frac{L - R}{L + R},$$

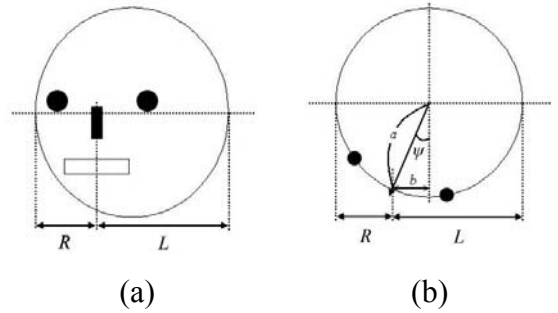$$\psi = \sin^{-1} \frac{L - R}{L + R}, \tag{8}$$

Fig. 3. Face and head model for estimating face direction angle. (a) Front view. (b) Top view.

IV.   Experimental Results

In our experiment, we test our system on images containing a static subject that have different direction. There are five persons with nineteen directions which a picture is taken from the front ranging from 90° to -90°. Table I shows the result of subject direction detection using the combined FD and LDA.Table II shows the result of face direction detection by the pupil locations.

We compared the direction estimation results of the methods which use subject direction detection and face direction detection.

7

Consequently, when the human direction angle (or face direction angle) is within $\pm 30^\circ$, the face is more accurately than that from the subject's silhouette. On the other hand, the face accuracy will decrease quickly when the angle is larger than $\pm 40^\circ$. Therefore, we can further fuse the face direction detection if the face is available and within $\pm 30^\circ$. Hence, we use the weights, in which w_1 and w_2, are constrained, as shown in Fig. 4., by the following relationship.
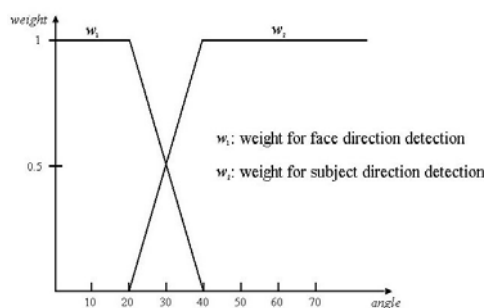

Fig. 4.   Weight graph of detection reliability.

The subject direction θ is estimated by using the weights and a linear interpolation method as follows.

$$\theta = w_1\psi + w_2\alpha, \qquad (9)$$

where ψ and α are the estimated face and subject direction angles, respectively. Table III is the estimated results by the proposed method which combine the results of subject and face direction detection.

To evaluate the estimation performance, we calculated the mean absolute error (MAE) of estimated angle of each model. Furthermore, we can estimate the average error of these five models. When subject direction detection using silhouette only, the average error is 4.58°. However, we fuse the face and subject direction detection, it can reduce 13% of error to 3.99°. As shown in Table IV, linear combination approach produces better performance than that from subject's silhouette only. Table IV shows the comparison of the MAE and average error of method 1: using only subject direction detection, and method 2: combine subject and face direction detection.

V. Conclusion

In this paper, we have proposed an approach which combine subject's silhouette matching and face direction for the human direction detection. In our direction detection system, we first utilize Fourier descriptor to discriminate different directions of target subject ranging from 90° to -90°. In addition, we exploit the linear discriminant analysis to optimize the class separability and improve the classification performance. Moreover, we also estimate the face direction from one's pupils and pupils' geometric relationship as well. When the human direction angle (or face direction angle) is within ± 30°, direction detection from the face is more accurately than that from subject's silhouette, and decrease quickly when the angle exceeds $\pm 40^\circ$. Therefore, we fuse to detect direction if face detection reliable. The experiment results have shown that our approach can obtain a high accuracy on subject direction detection, with mean absolute error less than 4°.

For future work, we shall increase the detection from the back subject's side ranging

from 90° to -90°. Therefore, we can complete the subject direction detection for all views. On the other hand, the distance of the subject is also important to be estimated, which is the subject of future research.

References

[1] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 809–830, August 2000.

[2] R. L. Cosgriff, "Identification of shape," Ohio State University Research Foundation, Columbus, *Report No. 820-11, ASTIA AD 254 792*, 1960.

[3] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," I*EEE Trans. Computers*, vol. C-21, pp. 269–281, 1972.

[4] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," in *Proc. First Int. Conf., AVBPA'97*, Crans-Montana, Switzerland, March 1997, pp. 127–142.

[5] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Trans. Multimedi*a, vol. 1, no. 3, pp. 264–277, September 1999.

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscienc*e, vol. 19, pp. 743–756, 1997.

[7] R. Feraud, O. J. Bernier, J. Viallet, and M. Collobert, "A fast and accurate face detector. based on neural networks," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 23, no. 1, January 2001.

[8] D. Maio and D. Maltoni, "Real-time face location on gray-scale static images," *Pattern Recognitio*n, vol. 33, pp. 1525–1539, 2000.

[9] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 21, pp. 557–563, 1999.

[10] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Trans. Intell. Transportation Syst.*, vol.7, no. 1, pp.63-77, Mar. 2006.

[11] P. Smith, P. M. Shah, and N. Lobo, "Determining driver visual attention with one camera," *IEEE Trans. Intell. Transportation Syst.,* vol. 4, no. 4, pp. 205- 218, Dec. 2003.

[12] Q. Ji , Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue, " *IEEE Trans. Vehicular Technology*, vol. 53, no. 4, pp. 1052- 1068, Jul. 2004.

TABLE I

THE ESTIMATED SUBJECT DIRECTION BY FD AND LDA

angel[°]	Model α [°]	Model α [°]	Model α [°]	Model α [°]	Model α [°]
90	85	83.56	85.30	83.61	86
80	74.65	76.33	84.53	82.98	75.9
70	65.73	74.4	65.67	75.61	74.12
60	55.13	63.62	64.24	65.10	56.48
50	53.43	46.24	54.35	54.80	54.07
40	34.18	37.24	43.86	43.61	43.47
30	26.05	26.3	26.31	34.45	26.26
20	15.16	16.2	23.70	23.84	15.97
10	5.74	5.98	6.41	13.57	5.97
0	-3.99	-3.02	-4.25	-3.33	-4.43

-10	-13.05	-5.97	-5.80	-16.78	-6.55
-20	-16.87	-15.87	-15.2	-26.43	-16.67
-30	-24.79	-26.68	-33.51	-35.88	-34.20
-40	-35.61	-43.59	-43.76	-44.37	-44.60
-50	-44.01	-54.02	-55.08	-55.17	-53.44
-60	-65.12	-63.9	-63.6	-63.89	-63.96
-70	-74.82	-65.76	-65.67	-85.81	-65.91
-80	-84.69	-85.92	-73.13	-83.99	-71.88
-90	-84.55	-86.78	-84.30	-85.20	-76.28

TABLE II

THE FACE DIRECTION BY THE PUPIL LOCATIONS

angel[°]	Model 1	Model 2	Model 3	Model 4	Model 5
	ψ [°]	ψ [°]	ψ [°]	ψ [°]	ψ [°]
60	12.33	54.23	36.57	27.54	-6.04
50	12.32	12.85	26-87	33.65	39.45
40	33.98	37.56	36-75	37.15	33.12
30	25.43	27.95	30.54	28.68	26.88
20	15.75	18.38	19.64	20.38	19.03
10	8	12.17	9.44	13.22	8.35
0	-2.86	-2.45	3.71	4.55	0.66
-10	-12.88	-15.97	-10.24	-15.65	-12.73
-20	-18.39	-22.84	-20.92	-25.14	-19.79
-30	-26.72	-30	-31.61	-33.54	-28.09
-40	-39.84	-36.78	-41.81	-29.78	-36.57
-50	-34.25	-31.69	-32.23	-23.17	-30.17
-60	-35.3	-24.86	-15.68	-24.7	-45.2

TABLE III

THE ESTIMATED SUBJECT DIRECTION BY THE PROPOSED METHOD

angel[°]	Model	Model	Model	Model	Model
	θ [°]	θ [°]	θ [°]	θ [°]	θ [°]
90	85	83.56	85.30	83.61	86
80	74.65	76.33	84.53	82.98	75.9
70	65.73	74.4	65.67	75.61	74.12
60	55.13	63.62	64.24	65.10	56.48
50	53.43	46.24	54.35	54.80	54.07
40	34.18	37.24	43.86	43.61	43.47
30	25.74	27.13	28.42	31.56	26.57
20	15.76	18.38	19.64	20.38	19.03
10	8	12.17	9.44	13.22	8.35
0	-2.86	-2.45	3.71	4.55	0.66
-10	-12.88	-15.97	-10.24	-15.65	-12.73
-20	-18.39	-22.84	-20.92	-25.14	-19.79
-30	-25.75	-28.34	-32.56	-34.7	-31.14
-40	-35.61	-43.59	-43.76	-44.37	-44.60
-50	-44.01	-54.02	-55.08	-55.17	-53.44
-60	-65.12	-63.9	-63.6	-63.89	-63.96
-70	-74.82	-65.76	-65.67	-85.81	-65.91
-80	-84.69	-85.92	-73.13	-83.99	-71.88

TABLE IV

MAE COMPARISON OF DIFFERENT METHODS

Method	Model	Model	Model	Model	Model	Average
Silhouette only	4.61°	4.37°	4.37°	5.31°	4.65°	4.58°
Combining	4.28°	3.64°	3.43°	4.87°	3.79°	3.99°