計畫名稱

臉部年齡辨識系統應用於人與機器人互動

摘要

本研究主題為開發人類臉部年齡辨識系統，並應用於人與機器互動。人臉隨著時間的推移不斷地在改變並且留下歲月的痕跡，我們選用全相貌影像(Appearance image)作為判斷年齡的特徵依據，並且將年齡分成共七個群組，每十歲為一組，大於60歲則皆屬於同一群組。以支持向量機(Support Vector Machine)來學習訓練大量影像資料庫，我們採用FG-NET和MORPH這兩個西方人臉資料庫，使電腦建立線性及非線性模型，最後結合F-measure的分數以加權的方式決定辨識出的所屬年齡群組。本研究方法可以實現在任何電腦或是機器人上，只需要配備有一台網路攝影機。

研究的問題所在及研究動機

以人為核心的社會環境當中，人與機器人的互動關係成為相當受到重視的一環，因此我們致力於開發各種人類資訊的偵測系統落實於電腦或機器人上，經由擷取人類的各種靜態與動態資訊，使得電腦或機器人有良好的先前知識來達到人機互動的目的，例如一個先進的健康照護系統，可以根據患者年齡自動配置適當的虛擬護士，隨時監控病情給予照顧，同時帶來互動的樂趣；另一方面藉由年齡資訊的擷取，我們能提升保全監控的應用面，例如管制未成年人從事違法行為等以降低人力資源的支出。

技術研究方法及創新性

方法流程如 Fig. 1。虛線左邊是人臉偵測和預處理的過程；虛線右邊是模型訓練和分類過程。首先說明訓練過程，所有輸入影像先分成男生和女生兩大類，接著根據真實年齡值，我們手動將訓練資料分成七個年齡群組：0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 以及大於等於 60 歲，所有影像透過支持向量機的學習得到線性和非線性兩種模型。

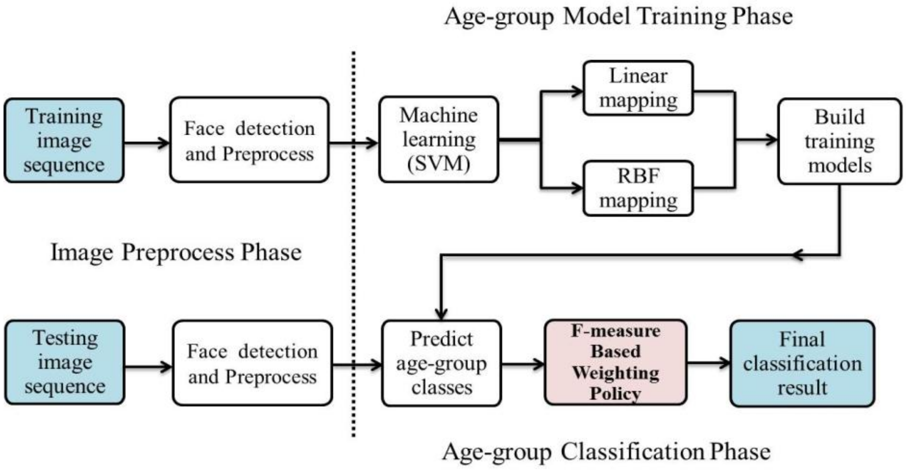

Fig. 1. Flow chart of age-group classification system.

於測試階段，所有輸入的測試影像根據前述得到的訓練模型，可以個別預測出其所屬年齡群組。

Confusion Matrices

機器學習領域的研究常常使用混淆矩陣來將結果可視化，我們從混淆矩陣中可以得到四個有意義的數值，介紹如下：

1) Accuracy: 所有 true positive 的樣本數除以樣本總數。

2) Recall: $R = \dfrac{\text{true positives}}{(\text{true positives} + \text{false negatives})}$

此值代表符合的樣本中被正確偵測出來的機率。

3) Precision: $P = \dfrac{\text{true positives}}{(\text{true positives} + \text{false positives})}$

此值代表找到的樣本中，符合我想要的機率。

4) F-measure: 在統計上的一種判斷標準，它同時考慮 Recall 和 Precision 的數值，在兩者之間取平均，公式如下，

$$F_\beta = \frac{\left(\beta^2+1\right)PR}{\beta^2 \cdot P + R}, \qquad \beta \ge 0$$

當 $\beta = 1$ 時，F-measure 的公式即變成調合平均數，$F_1 = \dfrac{2PR}{P+R}$ 。這個值介於 0 到 1 之間，數值愈大代表準確度愈高。

研究成果

(1)標準測試資料庫

我們使用 FG-NET 和 MORPH 兩個開放大型人臉資料庫，如圖 Fig. 2。

Fig. 2. Sample images of FG-NET and MORPH databases.

(2)實驗結果

■ 七群分類

實驗的影像資料統計如 Table 1，訓練和測試資料是隨機選擇並不重疊。實驗結果見 Table 2 和 Table 3，橫列是實際值而直行是預測結果，斜對角線稱做 Set-1，如女生是有紅底的框而男生是有藍底的框；若考慮一個群組的差異納入接受範圍，即是對角線加入淡黃色框的值稱做 Set-2。

Table 1. The number of images used in the experiment.

Experiment Database		C0	C1	C2	C3	C4	C5	C6	Total
Female	Train	87	96	97	98	103	65	45	591
	Test	28	29	32	36	34	14	13	186
Male	Train	144	114	111	104	111	97	49	730
	Test	39	32	33	33	34	20	14	205

Table 2. Confusion matrix of 7-class female age-group classification result.

Female	0-9	10-19	20-29	30-39	40-49	50-59	60up
	C0	C1	C2	C3	C4	C5	C6
C0	21	7	0	0	0	0	0
C1	5	23	1	0	0	0	0
C2	0	2	21	4	4	1	0
C3	0	0	1	35	0	0	0
C4	0	0	1	1	32	0	0
C5	0	0	0	2	3	9	0
C6	0	0	0	0	0	0	13

Table 3. Confusion matrix of 7-class male age-group classification result.

Male	0-9	10-19	20-29	30-39	40-49	50-59	60up
	0	1	2	3	4	5	6
0	34	4	1	0	0	0	0
1	4	24	3	1	0	0	0
2	0	9	17	7	0	0	0
3	0	0	3	25	4	1	0
4	0	0	0	1	32	1	0
5	0	0	0	1	2	17	0
6	0	0	0	0	0	3	11

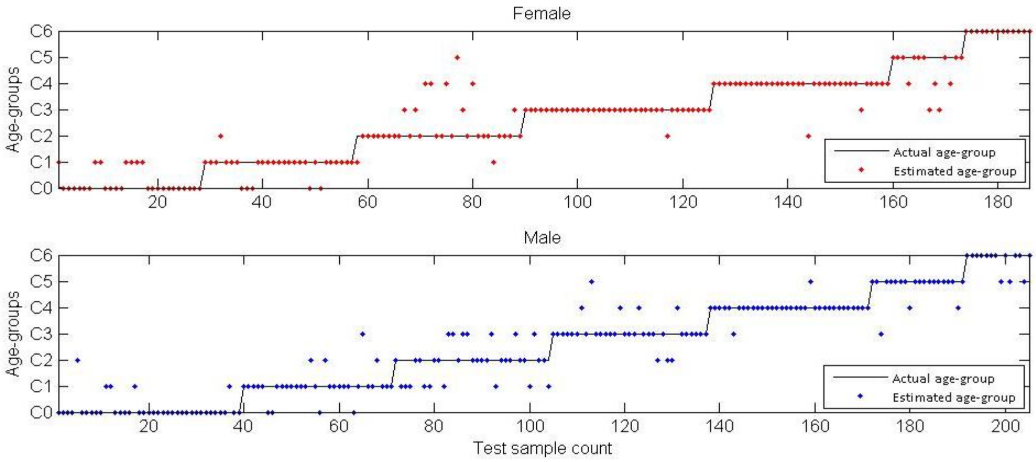

Fig. 3. Seven age-groups test samples classification results.

產業應用及其重要性

將系統以即時辨識的方式和電腦做互動，呈現結果如 Fig. 7。藉由人機介面的設計讓使用者簡單操作，達到人機互動的效果。

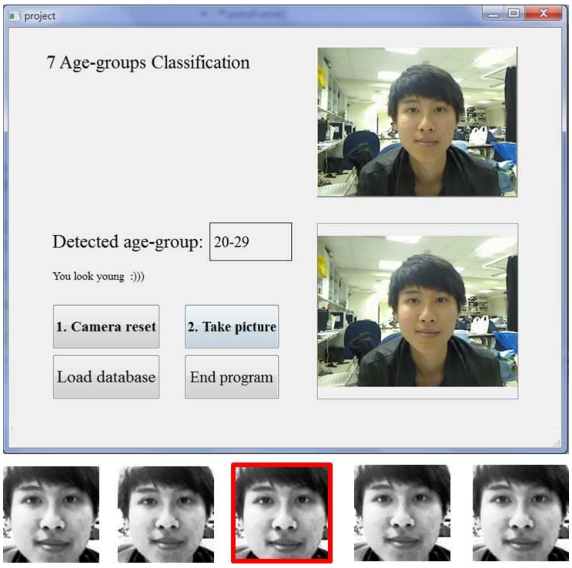

Fig. 4. A screen shot after clicking "Take picture" button. The result is 20-29 group shown on the user interface. A sequence of five continuous frames is listed in a row.

前面所提年齡辨識可以應用於保全監控功能,例如管制未成年人從事違法行為等,以販售香菸為例,利用年齡辨識系統杜絕未成年人購買煙品。

Tittle of Project

Age Estimation Using Appearance Images for Human-Robot Interaction

Abstract

There are many modern applications require the function of age estimation such as security control and surveillance monitoring, health care system and so on. In this study, we propose a method to classify human age using appearance images and apply it to the human-robot interactions. We first confirm that facial features based on craniology are not discriminative under the condition of seven age-groups classification. Next, our system is designed to have two stages. One is image preprocess stage; faces are detected and preprocessed. Our image database is from FG-NET and MORPH databases so that we have high degree of complexity in training dataset. Then images are trained by support vector machines (SVM). To have higher recognition rate, we train RBF (radial basis function) and linear kernel models at the same time, and decide the final results by F-measure based weighting policy. We also compare the age-group classification results with subjective questionnaires, and it demonstrates that the proposed system has better performance than human's subjective estimation. For the purpose of human-machine interaction, we design a simple user interface to perform online age-group classification. The system can be applied on any computer or robot as long as it has a camera sensor.

Problem Statement and Motivation

Since computer technology and information science change with each passing day, intelligent robots become more important in a variety of fields, such as industrial automation, military defense, security guard, in-home nurse, education and entertainment and so on. For this human-centered society, the interaction between human and robot is regarded as a significant part of the technological environment. Therefore, many scientists and researchers dedicated their time to develop all kinds of human information estimation systems so as to implement on computers and robots. Through the detection of static and dynamic human information, the purpose of human-machine interaction could be achieved according to the information.

Research Approach and Innovation

The proposed system is schematized in Fig. 1. For the face detection and preprocess are in the left to the dotted line. After preprocessing procedure, right to the dotted line is model training and classification phases.

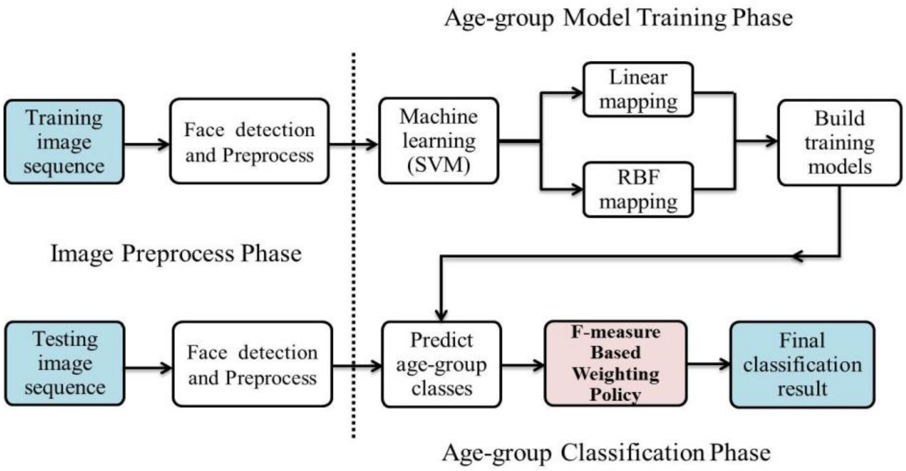

Fig. 5. Flow chart of age-group classification system.

The proposed system is schematized in Fig. 1. For the face detection and preprocess are in the left to the dotted line. After preprocessing procedure, right to the dotted line is model training and classification phases.

In training stage, all training data is firstly separated to two categories which are female and male. Given the ground truth age, we manually classify the training data into seven classes which are 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, and over 60 years old. Then, data is preprocessed and classified using support vector machines to build models. The system would produce two training models in the same time since we have two kernel functions: Linear and RBF.

After offline building training models, sequences of images are predicted and age-group results are acquired according to the two training models. When input a new face image, the system generates two possible results. By applying F-measure based weighting policy, the most possible answer is obtained to be the final classified age-group.

Confusion Matrices

Research of machine learning often uses confusion matrix to visualize the experimental results. We may obtain some useful information from the confusion matrix. There are four special terms explained as follows,

1) Accuracy: The sum of true positive values divided by the sum of total sample

counts. This is the most common representation of a learning machine's performance. Nevertheless, it is also easy to cover up the drawbacks of each class result solely. Therefore, we need other standard for further judgments.

2) Recall: Recall means how many samples are correctly detected. It is calculated using the following equation,

$$R = \frac{\text{true positives}}{(\text{true positives} + \text{false negatives})} \tag{1}$$

3) Precision: Precision represents the proportion of correct classifications versus samples recognized to be correct. Equation is written as,

$$P = \frac{\text{true positives}}{(\text{true positives} + \text{false positives})} \tag{2}$$

4) F-measure: It is an evaluation standard considering both recall and precision since sometimes these two values may have contradiction so either of them should not represent the true detection results. The F-measure definition is written as follows,

$$F_\beta = \frac{\left(\beta^2 + 1\right) PR}{\beta^2 \cdot P + R}, \qquad \beta \geq 0 \tag{3}$$

where β is a non-negative real value. Fβ is used to balance precision and recall so it could be seen as more fair information especially in critical conditions. When β = 1, Equation 3 is equivalent to the harmonic mean of P and R as shown in the following.

$$F_1 = \frac{2PR}{P + R} \tag{4}$$

This is called F1 or F1 score which averagely weighted the two mathematical values. It is a value between 0 and 1. If F1 is higher, that means the experimental method is more reliable.

Research Results

We collect two well-known databases, which are FG-NET and MORPH databases.

FG-NET:

MORPH:

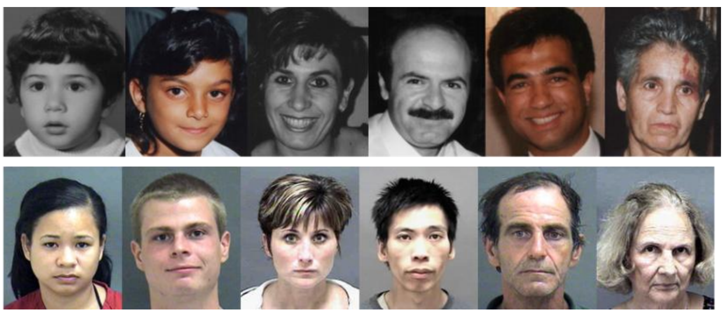

Fig. 6. Sample images of FG-NET and MORPH databases.

Table 4. The number of images used in the experiment.

Experiment Database		C0	C1	C2	C3	C4	C5	C6	Total
Female	Train	87	96	97	98	103	65	45	591
	Test	28	29	32	36	34	14	13	186
Male	Train	144	114	111	104	111	97	49	730
	Test	39	32	33	33	34	20	14	205

◆Seven age-group classification:

錯誤! 找不到參照來源。 shows the number of images used in the experiment. C0 to C6 represent the seven age groups from youth to elder. The training and testing images are randomly chosen from two databases without overlapping.

The classification results of female and male are shown in Table 2 and Table 3, respectively. The most left column represents ground truth age-group and the second row from top means the detected age-group. We consider two sets, Set-1 and Set-2. In Table 2, set-1 labeled with red color implies exactly correct classification and the average recognition rates of Set-1 reach 82.95%. For male result in Table 3, the recognition rate of Set-1 labeled with blue color achieves to 78.49%. In Fig. 5-4, detailed classification results are plotted. If red and blue dots are not on the black line, they are wrong classification samples.

Table 5. Confusion matrix of 7-class female age-group classification result.

Female	**0-9**	**10-19**	**20-29**	**30-39**	**40-49**	**50-59**	**60up**
	C0	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**
C0	**21**	**7**	**0**	**0**	**0**	**0**	**0**
C1	**5**	**23**	**1**	**0**	**0**	**0**	**0**
C2	**0**	**2**	**21**	**4**	**4**	**1**	**0**
C3	**0**	**0**	**1**	**35**	**0**	**0**	**0**
C4	**0**	**0**	**1**	**1**	**32**	**0**	**0**
C5	**0**	**0**	**0**	**2**	**3**	**9**	**0**
C6	**0**	**0**	**0**	**0**	**0**	**0**	**13**

Table 6. Confusion matrix of 7-class male age-group classification result.

Male	0-9	10-19	20-29	30-39	40-49	50-59	60up
	0	1	2	3	4	5	6
0	34	4	1	0	0	0	0
1	4	24	3	1	0	0	0
2	0	9	17	7	0	0	0
3	0	0	3	25	4	1	0
4	0	0	0	1	32	1	0
5	0	0	0	1	2	17	0
6	0	0	0	0	0	3	11

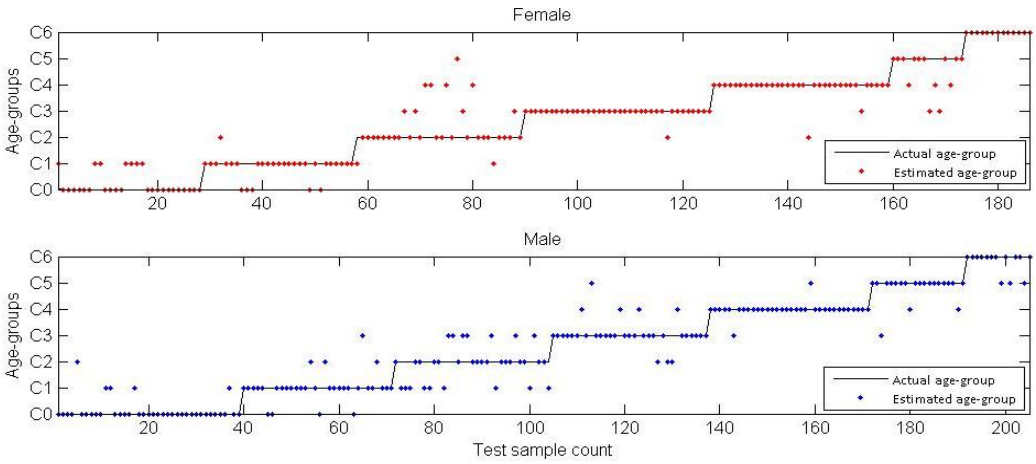

Fig. 7. Seven age-groups test samples classification results.

If we take one group of difference into consideration, then it forms Set-2 of the exact classified age-group plus the neighbored age-groups. For female, the red label plus skin color label are Set-2 range. For male, the blue label plus skin color label are Set-2 range. The recognition rates of Set-2 increase to 95.7% for female and 98.05% for male.

Industrial Applications and Its Impact

In our experiments, we design a graphical user interface that allow user to interact with robot or computer in real-time. In 錯誤! 找不到參照來源。, we invited a male student whose actual age is 23 years old for testing. The result is correctly classified to the "20-29" age-group and a comment "You look young" was also shown on the user interface.

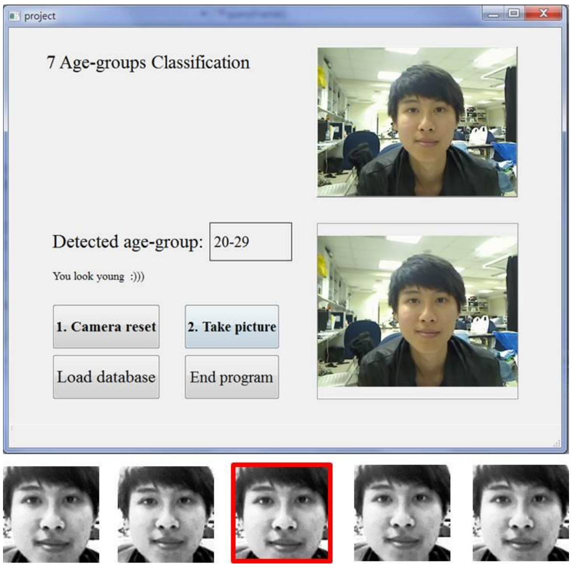

Fig. 8. A screen shot after clicking "Take picture" button. The result is 20-29 group shown on the user interface. A sequence of five continuous frames is listed in a row.

Our proposed system is designed to help improve functions on service robots with safety and health care in mind. In the case of selling cigarettes, robot has the ability to recognize minors and refuse their purchasing.