

行政院國家科學委員會專題研究計畫 期中進度報告

子計畫二：人類動作、情緒與生理訊號監控系統研發及其於 健康狀態偵測與維護之應用(1/2)

計畫類別：整合型計畫

計畫編號：NSC94-2213-E-009-097-

執行期間：94年08月01日至95年07月31日

執行單位：國立交通大學電機與控制工程學系(所)

計畫主持人：張志永

報告類型：精簡報告

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：本計畫可公開查詢

中 華 民 國 95 年 6 月 1 日

行政院國家科學委員會補助專題研究計畫 第一年期中成果報告

總計畫：以生理訊號為基礎之人機介面設計與應用
—人類操控機器的新模式

子計畫二：人類情緒與生理訊號監控系統研發及其於健康
狀態偵測與維護之應用(1/2)

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC-94-2213-E-009-097

執行期間：94年8月1日至95年7月31日

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

執行單位：國立交通大學電機與控制工程學系

中華民國 95 年 5 月 30 日

行政院國家科學委員會專題研究計畫 第一年期中期報告

**總計畫：以生理訊號為基礎之人機介面設計與應用
—人類操控機器的新模式**

**子計畫二：人類情緒與生理訊號監控系統研發及其於健康
狀態偵測與維護之應用(1/2)**

**(Subproject II) A Study of Emotional and physiological Signal Monitoring
System for Intelligent Health Care Application (I)**

計畫編號：NSC-94-2213-E-009-097

執行期限：94/08/01-95/07/31

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

本計畫研發一個以影像為基礎的人類動作、表情情緒與生理訊號監控系統，並使用機器學習技術，由所監控影像之動作、表情與生理訊號共同建構一個健康狀態辨認知識系統[1-4]，以驗證其於健康狀態自動偵測與健康維護應用之可行性。本年度研發重點如下：藉由彩色 CCD 攝影機拍攝家裏人的運動情況，並能自動分析他的動作；亦即本報告提出一個能夠自動監控並且辨識人類動作的初步研究。我們利用資料轉換將影像轉換至特徵空間(Eigenspace transformation)，再轉至標準空間(Canonical space transformation)，提供一個能夠在視訊影像中辨識人類動作的系統。在我們的系統中，每一影像序列的前景人物會利用一個背景模型而被抽取出來，並以二值化影像代替。背景模型會使用到連續影像的相除(frame

ratio)。接著，二值化影像經由特徵空間及標準空間轉換投影至標準空間。最後人類動作的識別在標準空間中完成。我們所提供的系統對每一類的動作僅使用幾個必要且有效的樣板來代表，而不使用全部的影像序列；也就是對影像序列作降頻取樣，這麼做的好處是可以降低辨識問題的複雜度、減低運算負載並且增加辨識率。我們提出的這個系統僅僅使用這些二值化的影像來辨識人物的動作，而且沒有參考其他任何資訊例如位置、路徑或速度等等，並有相當高的辨識率。

關鍵詞：動作辨識、模糊辨識器、特徵選擇與轉換、健康監控系統

英文摘要

It is known that human activity, facial expression, and physiological signals explicitly

shed light on the health and comfort status of a person. Based on recognizing video-based human activity, facial expression, and physiological signals, the purpose of this project is to build an automated health monitoring system to determine the physical and mental comfort of a person so as to predict one's health condition [1-4]. With an objective signal conditioning, we will develop a machine's ability to recognize human affective health state by watching through a CCD camera over a person's action and face. The objective of this thesis is to provide a human-like system to auto-surveillance, to track people and to identify their activities. We present a system for video-based human activity recognition by transforming the images into canonical space. In our system, foreground subject is first extracted as the binary image by a statistical background model using frame ratio, which is robust to illumination change. Then the foreground binary image is transformed by eigenspace and canonical space transformation, and recognition is done in canonical space. By using several essential templates to represent an activity, our proposed system can recognize the activity of the subject by down sampling the image sequence instead of all consecutive image frames. In this way, we can reduce the recognition complexity, decrease the computational load, and improve the recognition performance. Without referring to any geographic information such as location, path, and velocity of the subject, our proposed system uses only the binary images of subject to recognize the activity and works very well.

Keywords: Action recognition, fuzzy classifier,

feature selection and transformation.

二、緣由與目的

本計畫將研發一個人類動作、情緒與生理訊號監控系統，並使用機器學習技術，由監控影像與生理訊號建構一個健康狀態辨認知識系統系統，以驗證其於健康狀態偵測與維護應用之可行性。利用子計畫三對腦電波和感覺知覺基本研究，如注意力、疲倦等與腦電波之關係，可探討發現人健康狀態最強而有力原始癥兆，其研究成果可成為本計畫二最直接有效訊號；而子計畫四、六是血壓計與腦電波量測儀器和重要零組件研製，其研究成果，即是本計畫最實用之量測儀器，並極有助於非察覺式之量測儀器裝置佈建，增進本計畫應用之可行性。子計畫一可以對目前與未來短暫時間內人體之狀態與行為進行判斷與預測，預測的結果即可回饋給本計畫的控制電腦；本計畫成果將運用於醫療看護系統，使得數位計算器更「自動瞭解」人類之身心舒適與健康狀態之行為模式，進而「自動產生」對應的操作或醫療模式 [1, 2]，使得人機介面達到人性化的目標。

從串流視訊影像資料中識別人類動作有相當多應用，例如人機介面、安全監控系統、居家看護等等。近來，有相當多類似的題目及方法。Yamato 等人 [9] 利用 Hidden Markov Model (HMM) 結合時間序列影像來辨識人類動作。Bobick 和 Davis [10] 利用暫時性樣板結合 motion-energy images 和 motion-history images 來識別人類動作。Bodor 等人 [11] 利用行人行走的速度和路徑來辨識公共空間中人類的動作。Masoud 和 Papanikolopoulos [12] 則是利用主成分分析在轉換空間中識別人類動作。

在偵測、追蹤前景人物及識別前景人物

動作上，有已完成且具參考價值的計畫。而 W^4 [8], [13] 即是其中的一員。 W^4 利用一個具環境適應性的背景模型將前景人物從背景中分離出來，並使用前景人物的外型輪廓來辨識人物的動作。在 [14]–[17] 中，他們將前景人物的姿勢利用一組經標準化的肢體角度來表示。因此，人類動作就可以用一組多維度的向量來表示，而這向量代表的意思即是在整個動作週期內被取樣道的肢體角度以及角速度。利用這些向量即可建立一個 Hash table，每一組 Hash table 表示一個不同的肢體。

有許多的資料轉換方法被用在縮減資料的維度。Huang et al. [3] 將特徵空間轉換 (Eigenspace transformation) 及標準空間轉換 (Canonical space transformation) 應用在人臉及人類步法識別上。另外，Jobson and Woodell [4], and Rahman [5] 建立了一套影像增強的理論稱為 retinex。雖然我們的實驗中並不需要增強所使用的影像，但是 retinex 所使用的理論引導我們利用影像相除 (Frame Ratio) 建立所需的背景模型以及前景人物抽取。在前景人物抽取方面，Park 及 Aggarwal [6] 利用 HSV 色彩模型計算每一個像素的 Mahalanobis distance 來分離前景及背景。Leung and Yang [7] 建立了一個人體輪廓標記系統，而 Haritaoglu 等人 [8] 則是建立了人體肢節標記系統。這兩份論文皆試圖找出人在影像上的真正姿勢。

本報告提供一個能夠自動監控、追蹤並且辨識人類動作的系統。我們所提出的系統利用灰階影像來做追蹤以及辨識能找出前景人物在整張影像上的位置，並且辨識出前景人物的動作。

三、研究方法

(1) 特徵空間及標準空間轉換

在串流視訊影像處理上，影像序列的資料量通常極端的大。有許多相當知名的資料轉換方法可以用來縮減資料維度，例如 PCA，DCT，小波轉換等。在我們的系統中，我們利用特徵空間轉換來縮減影像序列的維度。在動作識別方面，影像序列的比對是直得進一步研究的關鍵步驟。在我們的識別系統中，我們利用標準空間轉換來增加識別率。基於主成分分析的特徵空間轉換對於縮減資料量是個強而有力的公具，在縮減資料量的同時，也能保留資料的代表性。而基於標準分析的標準空間轉換能最佳化類別間的分散度，進而增加分類表現。人物動作的識別即是在標準空間中完成的。 S_w 表示同類別中向量距離的平均， S_b 表示類別間向量距離的平均。我們的目標就是同時地將 S_b 最大化且將 S_w 最小化。而這個目標就是 generalized Fisher linear discriminant function：

$$J(W) = \frac{W^T S_b W}{W^T S_w W}. \quad (1)$$

藉由可選擇的 W 可將上述之變異量的比值最大化，即選擇 W 使得

$$\frac{\partial J}{\partial W} = 0.$$

(2) 影像序列的前處理

在整個人類動作識別系統中的第一步，即是建立背景模型。在本論文中，我們將背景模型描述為可運算的統計模型。因此，需要一段背景視訊影片來統計，並同時紀錄每一個像素上的最大及最小的灰階值、和最大的 frame ratio。

我們發展了一個對光線變化有適應性的方法，這個方法稱為 frame ratio。就效果來

說，使用 frame ratio 會較傳統的 frame difference 來的有效。通常，同一個地點的光線變化是緩慢且柔和的，但是長時間的光線變化依然會影響整個背景模型。在本報告中，我們利用三個統計值來描述背景模型：每個像素上的最大灰階值 $m(x, y)$ ，每個像素上的最小灰階值 $n(x, y)$ ，以及最大的 interframe ratio $d(x, y)$ ，根據我們所建立的背景判定模型，每一張影像序列上的前景人物皆可被分離出來。利用這個背景模型，每一個像素都會有一個是不是背景像素的分類結果，並建立出一張二值化的影像。 **圖一

(3) 動作樣板的選擇

我們選擇某一個動作影像序列中的某幾張影像當成這個動作的代表樣板。這種動作樣板的表示對於減少運算量有相當大的幫助，因為在辨識階段的比對不需要這類動作的全部影像的比對，取而代之的是比對我們選擇的有效樣板。以少數的樣板來代表某一類動作的依據是根據我們對動作的觀察，在連續的動作影像序列中，鄰近的影像上的前景人物的姿勢是非常接近的，我們預期在經由資料轉換後，這些相似的動作會被分類至同一個有效樣板。在實驗中，我們設計了四類的動作，分別是“由右朝左走”、“由左朝右走”、“跳”、“蹲下”。我們對“右朝左走”與“由左朝右走”分別選擇九個有效樣板，對“跳”及“蹲下”分別選擇一個有效樣板，對“共同狀態”選擇三個有效樣板。“共同狀態”指的是跳與蹲下時會有的共同姿勢。圖三顯示共同狀態與跳跟蹲下的關係。在本論文中，僅僅使用這 23 個有效樣板（也就是共有 23 種類別）來識別這四種動作。

(4) 分類法

每一二值化序列影像在分類之前，都會

先由先前所建立的轉換矩陣 H 將影像投影至標準空間中。系統會將轉換至標準空間的二值化影像與之前所訓練完成的訓練組 Z 作比對，並作出前景人物動作的識別。我們所提出的系統用了兩種分類法：nearest neighbor 以及 maximum likelihood 分類法。甲、nearest neighbor 分類法：給定一未知類別的 t_k ，系統會將 t_k 與訓練組 Z 內所有的行向量 $z_{i,j}$ 計算距離，並同時將 t_k 指定為與最近的 $z_{i,j}$ 同為第 i 類，其中， $z_{i,j}$ 表示第 i 類中的第 j 個行向量。而此處所指的距離是在標準空間中的 Euclidean distance。乙、maximum likelihood 分類法：假設 C_i 為第 i 類， C_i 為第 i 類中的 covariance matrix，則 likelihood function 可表示為

$$\begin{aligned} L(t_k|C_i) &= \frac{1}{(2\pi)^{\frac{c-1}{2}} \det^2(C_i)} \exp\left[-\frac{1}{2} t_k^T C_i t_k\right] \\ &= \prod_{m=1}^{c-1} \frac{1}{\sqrt{2\pi}\sigma_{i,m}} \exp\left[-\frac{1}{2} \frac{(t_k^m - \mu_{i,m})^2}{\sigma_{i,m}^2}\right], \end{aligned}$$

則系統將一未知類別的 t_k 指定至第 p 類根據：

$$p = \arg \max_i L(t_k|C_i)$$

四、實驗與結果

在我們的實驗當中，先前所選擇的有效樣板即是我們的訓練組，首先這些有效樣板(訓練組)會先經過特徵空間轉換與標準空間轉換，並建立出所需的轉換矩陣 H 。接著，每一個視訊影像序列中的前景人物皆會經由背景模型抽取二值化前景影像。被取樣到的二值化影像接著利用轉換矩陣 H 投影至標準空間中。最後利用兩個分類法，nearest neighbor 與 maximum likelihood 分稱為演算法一與二，識別前景人物

的動作。辨識正確率的比較分為兩序列影像中的每一張影像皆取樣出來，並比較之。在有硬體實現限制的演算法的部分，由於對視訊序列影像作了降頻取樣，因此有硬體實現限制的演算法只對彼此作比較。藉由觀察“由右朝左走”及“由左向右走”，我們發現所選擇的樣板之間的時間間隔大約是五個攝影機取樣時間(5 frames)，因此，針對兩個有硬體實現限制的演算法，我們假設執行速度是每秒六張影像，也就是 5:1 的降頻取樣。在對影像序列作比對時，一個常碰到的問題是比對序列起始點不同會影響比對的結果。根據我們的假設，我們比較了五個不同的起始點，並將結果列在表中。正確率的計算是將被取樣且分類正確的影像數目除已被取樣的影像總數。在實驗中，我們使用 leave one out cross validation 測試系統，意即取其中一人的序列影像當成測試影像，其餘的人的序列影像即為訓練影像。實驗中所使用的影像大小皆為 640×480 的八位元灰階影像，而被抽取出的二值化前景影像大小皆為 128×96 二值化影像。

實驗中，共有七個人，這七個人分別做了四種動作：“由右朝左走”、“由左朝右走”、“跳”、“蹲”。其中，跳與蹲會包含共有的姿勢稱為“共同狀態”。因此，實驗中我們利用五個大分類來代表四種動作，而總類別共有 23 類(23 個有效樣板)。對於 nearest neighbor 與 maximum likelihood 分類法的正確率比較中，總共有 2871 張影像。對於演算法一與二，我們假設執行速度是每秒六張影像，也就是 5:1 的降頻取樣。兩種分類法的比較將列於表一中，而兩種演算法的比較列於表二與表三。

(1) 利用 Fuzzy Rule 辨識人類動作

時間序列的資訊在人類動作辨識上，是一個

重要的判斷指標。我們透過 fuzzy rules 來學習人類做動作時的姿態轉換模組，利用此模組來做為動作識別的依據。Fuzzy rules 具有容忍模擬兩可資料的特性，也可以學習出在我們定義姿態轉換模組外的隱藏轉換模組，藉此可以提升系統的辨識率。Wang 等 [18]提出了 fuzzy rules 可以透過範例的學習來建立。在我們的系統中做學習的時候，首先將挑選出代表各動作的樣板，接著將我們的訓練影像序列輸入系統中學習，系統會先將影像透過 EST 和 CST 的轉換抽取影像中的訊息，利用轉換後的結果計算出該影像相對於各樣板的 membership functions 的值。我們利用三個影像為一組，然後計算該組合中所有可能的 membership functions 的總合大小，將值最大的組合符號序列透過 IF-THEN 的形式建立到資料庫中。當中的符號序列就是利用樣板的代號來做為動作變化的表示，所以經由此方法訓練學習可以將每一個動作中可能出現的轉換形式學到 fuzzy rules 的資料庫中。

當系統做動作識別的時候，相同的我們先將影像利用 EST 及 CST 做轉換，也是以三個影像為一組，透過空間轉換後的向量值計算出其相對於各樣板的 membership functions 值，接著系統會去資料庫中比對各個 rule。比對的時候，會依照 IF-THEN 中的紀錄符號序列去取出影像組合中相對應位置影像的 membership functions 值，然後加總起來。當最後將每一個 rule 的總合結果都計算出來後，選取所有 rule 中 membership functions 總合最高分的 rule，並依照 rule 訓練時所紀錄動作，將未知的影像序列歸類為此項動作。在此方面，我們這半年正在進行上一年度影像序列，動作分類的正確率有 91.8%，約比最常用之 Hidden Markov Model 提高 5.4%。

五、結論與討論

人類動作的辨識在相當多的領域中有相關的應用，例如保安系統、居家看護等。本報告中，我們提出了一個在串流視訊影像資料中，將影像轉換至標準空間進而識別人類動作的系統。在我們所提出的系統中，被取樣的前景影像會先利用背景模型將前景抽取出來，並以二值化影像表示，接著利用特徵空間與標準空間轉換將影像投影至標準空間，並在標準空間中完成動作的識別。我們所提出的系統完全不參考任何前景人物在影像中的地理資訊(位置、路徑、行進速度等)，而僅僅利用二值化前景影像就能將影像上的人物所表現的動作辨識出來，並且有著相當高的辨識率。我們利用少數的有效樣板來代表一個動作，而這種表示方式經我們的實驗證實的確能夠充分表示，且在比對的階段不需要全部的影像序列即可作出辨識。這個表示方式有效的降低了辨識問題的複雜度，同時降低運算負載，並且增加辨識率。

總結來說，我們使用一個能夠正確描述所設計的四種動作的轉換方法，且在辨識問題上提出了一個強健的系統。而這個系統能有效率的減低資料維度且對光線變化有良好的適應性。本計畫已有好的開始與成果，下一年度將進行更精緻之動作分類，及健康異樣時之動作差異與分辨等研究。

六、參考文獻

- [1] 「老人照護產業 資訊通訊科技應用及商機」國際研討會，經濟部資訊工業發展推動小組，2004年10月，台北市。
- [2] D. C. Lewis, "Predicting the future of health care," *The Brown University Digest of Addiction Theory & Application*, vol. 18, no. 4, pp. 12-16, 1999.
- [3] P. S. Huang, C. J. Harris, and M. S. Nixon, "Human gait recognition in canonical

space using temporal templates," *Vis. Imag. Signal Process.*, vol. 146, no. 2, pp. 93-100, 1999.

- [4] D. J. Jobson and G. A. Woodell, "Properties of a center/surround retinex part two: surround design", *NASA Technical Memorandum #110188*, 1995.
- [5] Z. Rahman, "Properties of a center/surround retinex part one: signal processing design", *NASA Contractor Report #198194*, 1995.
- [6] S. Park, J. K. Aggarwal, "Segmentation and Tracking of Interacting Human Body Parts under Occlusion and Shadowing," in *Proc. of the Workshop on Motion and Video Computing*, pp.105, Dec. 05-06, 2002.
- [7] M. K. Leung and Y. H. Yang, "First sight: A human-body outline labeling system," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 4, pp. 359-377, Apr. 1995.
- [8] I. Haritaoglu, D. Harwood, and L. Davis, "Ghost: A Human Body Part Labeling System Using Silhouettes," in *Proc. Int'l Conf. Pattern Recognition*, 1998.
- [9] J. Yamato, J. Ohya, K. Ishii, "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," In *Proc. IEEE CVPR*, pp. 379-385, 1992.
- [10] A. F. Bobick and J. W. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 3, Mar. 2001.
- [11] R. Bodor, B. Jackson and N. Papanikolopoulos, "Vision-Based Human Tracking and Activity Recognition," in

Proc. of the 11th Mediterranean Conf. On Control and Automation, June 18–20, 2003.

- [12] O. Masoud and N. Papanikolopoulos, "Recognizing human activities," in *Proc. IEEE Conf. Advanced Video and Signal Based Surveillance*, Miami, FL, Jul. 2003, pp. 157–162.
- [13] I. Haritaoglu, D. Harwood and L. S. Davis, "W4: Real-Time Surveillance of People and Their Activities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.22, no. 8, pp. 809–830, August 2000.
- [14] J. Ben-Arie, Z. Wang, P. Pandit, and S. Rajaram, "Human Activity Recognition Using Multidimensional Indexing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp.1091–1105, August 2002.
- [15] K.R. Rao and J. Ben-Arie, "Multiple Template Matching Using the Expansion Filter," *IEEE Trans. Video Technology*, vol. 4, no. 5, pp. 490–504, 1994.
- [16] K R Rao and J. Ben-Arie , "Optimal Edge Detection Using Expansion Matchingand Restoration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13,no. 12, pp. 1169–1182,Dec. 1994.
- [17] J. Ben-Arie and KR Rao , "A Novel Approach for Template Matching by Nonorthogonal Image Expansion," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 3,no. 1, pp. 71–84, 1993.
- [18] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples", *IEEE Trans. Syst. Man Cybern*, Vol. 22, No. 6, PP. 1414-1427,

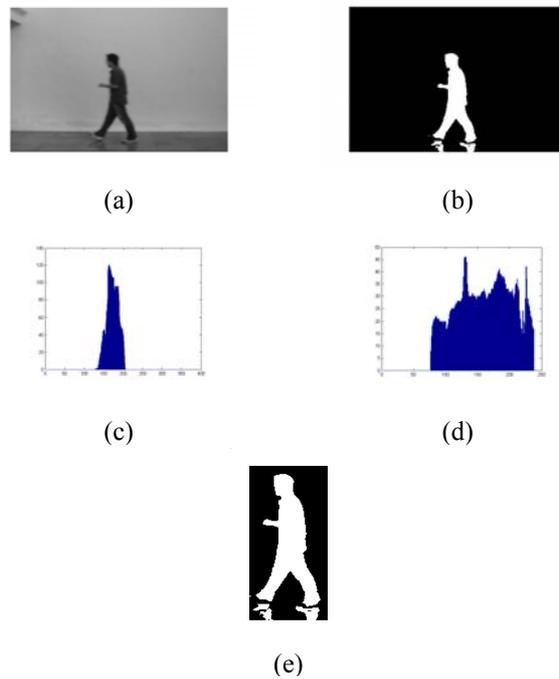


圖 一：一個前景影像抽區的範例。(a) 某一影像序列，(b)經由背景模型分析後所得到的二值化影像，(c)二值化影像在垂直方向的投影，(d)二值化影像水平方向的投影，(e)最後所得到的前景二值化影像。

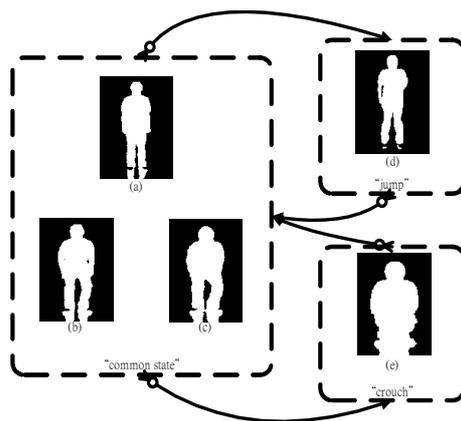


圖 二：跳(d)和蹲(e)皆會有的共同姿勢(a)、(b)、(c)。

表一 Nearest neighbor 與 maximum likelihood 兩種分類法的正確率

	Nearest neighbor	Maximum likelihood
Person 1	87.6	90.3
Person 2	77.6	89.3
Person 3	92.9	93.9
Person 4	92.0	89.5
Person 5	93.0	94.7
Person 6	78.1	80.2
Person 7	93.3	96.4
Average	87.8	90.2

表二 演算法一的正確率。

	Start at sample 1, i.e., 1, 6, 11, 16, 21, ...	Start at sample 2, i.e., 2, 7, 12, 17, 22, ...	Start at sample 3, i.e., 3, 8, 13, 18, 23, ...	Start at sample 4, i.e., 4, 9, 14, 19, 24, ...	Start at sample 5, i.e., 5, 10, 15, 20, 25, ...
Person 1	92.7	96.3	100.0	100.0	96
Person 2	87.5	93.7	85.7	88.7	86.9
Person 3	98.4	96.9	100.0	96.8	95.1
Person 4	98.6	98.6	95.8	97.2	98.6
Person 5	100.0	100.0	97.4	98.7	98.7
Person 6	86.3	87.7	89.0	83.1	78.6
Person 7	100.0	100.0	95.1	95.1	100
Average	94.9	96.1	94.6	94.1	93.3

表三 演算法二的正確率

	Start at sample 1 i.e., 1, 6, 11, 16, 21, ...	Start at sample 2 i.e., 2, 7, 12, 17, 22, ...	Start at sample 3 i.e., 3, 8, 13, 18, 23, ...	Start at sample 4 i.e., 4, 9, 14, 19, 24, ...	Start at sample 5 i.e., 5, 10, 15, 20, 25, ...
Person 1	96.4	94.4	94.4	98.1	98
Person 2	98.4	98.4	96.8	95.2	100
Person 3	100.0	100.0	100.0	100.0	100
Person 4	100.0	100.0	90.1	85.9	92.9
Person 5	100.0	100.0	100.0	100.0	100
Person 6	89.0	93.2	91.8	87.3	85.7
Person 7	100.0	100.0	100.0	98.4	100
Average	97.7	98.1	96.1	94.8	96.5