行政院國家科學委員會專題研究計畫 成果報告

基於延展式區域三元化樣式的物件描述方法
研究成果報告(精簡版)

計 畫 主 持 人 ：廖文宏

計畫參與人員：博士班研究生-兼任助理人員：林明慶

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 100 年 10 月 30 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

基於延展式區域三元化樣式的物件描述方法

計畫類別：■個別型計畫　　□整合型計畫
計畫編號：NSC 99－2221－E－004－012－
執行期間：　99　年 8 月 1 日至 100　年 7 月 31　日

執行機構及系所：國立政治大學資訊科學系

計畫主持人：廖文宏
共同主持人：
計畫參與人員：林明慶

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

本計畫除繳交成果報告外，另須繳交以下出國心得報告：
■赴國外出差或研習心得報告
□赴大陸地區出差或研習心得報告
■出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

中　華　民　國　100　年　10　月　30　日

基於延展式區域三元化樣式的物件描述方法

摘要

特徵描述子為電腦視覺中相當重要的一部分，本研究基於知名的特徵描述子：區域二元化圖型的架構上，提出了新的特徵描述子，並將其命名為延展式區域三元化圖型。我們所提出的特徵描述子與區域二元化圖型相比，有著較強的抗噪能力而且保留了區域二元化圖型簡單的計算複雜度。本計畫也探討了區域三元化圖型中是否存在著 uniform pattern，基於區域二元化圖型中 uniform pattern 的定義，我們提出了屬於區域三元化圖型的 uniform pattern，並在圖像實驗中驗證了其大量存在性。我們將區域三元化圖型應用於材質分析與人臉辨識中，實驗結果驗證了本計畫所提出的特徵描述法在這些應用的優越性。

關鍵字：區域二元化圖型、延展式區域三元化圖型、uniform pattern、材質辨識

Object Description Using Extended Local Ternary Pattern

Abstract

Robust feature descriptor is essential in developing effective computer vision applications. In this report, we present an extension to the well-known local binary pattern (LBP) feature descriptor. The newly defined descriptor known as extended local ternary pattern (ELTP) exhibits better noise resistivity than the original LBP, while maintaining computational simplicity. We further investigate the presence of uniform patterns in ELTP. With a slight modification of the definition of uniformity, it is found experimentally that uniform ELTPs account for 80% of all patterns in texture images. The proposed ELTP and uniform ELTP are applied to object classification tasks, including texture analysis and face recognition. Experimental results validate the superiority of ELTP over conventional LBP approaches.

Keywords：local binary pattern, extended local ternary patterns, uniform pattern, texture classification

基於延展式區域三元化樣式的物件描述方法

1. 研究背景及目的

    物件辨識(object recognition)向來是電腦視覺領域中的重要課題，準確而有效率的物件辨識演算法，除可作為機器人視覺的伺服控制機制，在視覺搜尋(visual search)、分類(visual object class，VOC)、多媒體內容分析、安全監控也都有直接的應用。然而物件辨識是一項具挑戰性的任務，除了外在環境變因的不可控制性，影響影像分析與特徵擷取的穩定度外，使用視覺資訊，還有一項最大問題，就是如何從龐大的資料中進行有效率的運算，找出所需追蹤或辨識的特徵，進而獲得辨識或理解週遭事物的能力。英語中有句諺語是這樣說的:「A picture is worth a thousand words」，意謂一張圖像可以勝過千言萬語的描述，然而這是基於人類的認知與理解能力所引申的觀察，從電腦的角度而言，一張未經壓縮的百萬畫數的全彩相片(1 million pixels，24-bit color)需要 3 百萬個位元組(byte)來表示，也就是: "A picture is worth a thousand 'words'. Yet it takes millions of 'bytes' to store a picture."

    而相片內容的分析與處理，通常需要耗費更多的計算資源，但卻很難達到令人滿意的效果，也就是說，從計算的角度而言，使用視覺的描述方式雖然看似直覺，但卻不一定是最有效率的。近來由於電腦硬體技術的大幅進展與實際需求面的增加，特定類別的物件辨識，如人臉偵測(face detection)[1]、行人偵測(pedestrian detection)[2]、文字辨識(text recognition)[3]等，有了長足的進步，部分應用甚至已經從個人電腦轉至智慧型行動平台，增加了便利性與實用性，例如最近 Google 推出的 Google Goggles 服務[4]，就是利用手機拍攝的照片，進行圖像搜尋，透過其強大的雲端運算後台，有效率地回傳與圖片內容相關的訊息，當然該項服務還屬於實驗階段，並非十全十美，而且僅針對特定物件類別，如著名景點、藝術品、名片、書的封面、商標等進行最佳化處理，在物件辨識的準確率與通用性方面，仍有改善的空間。

    在物件類別辨識的研究方面，歐盟所支援的 PASCAL (Network of Excellence on Pattern Analysis, Statistical Modeling and Computational Learning)從 2005 年開始就舉辦 Visual Objects Classes (VOC) Challenge，累積至 2010 年已建立二十項物件資料庫[5]，包含

- Person: person
- Animal: bird, cat, cow, dog, horse, sheep
- Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
- Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor

    該競賽每年均吸引數十個來自世界各地的研究團隊參與，並於ICCV舉辦workshop發表比賽結果，以 2010 年最新資料顯示，各類物件辨識的平均準確率(average precision)從 30-80% 不等，顯示在這方面的研究也還有精進的空間。

    一個強健的物件辨識核心，通常需具備幾個要素，包含可靠的特徵描述方式(feature descriptor)，以及兼具效能與正確性的分類與辨識演算法。在物件描述方面，近來常被使用的影像特徵有 Harr-like features[1]、Scale-Invariant Feature Transform (SIFT)[6]、Histogram of Oriented Gradients (HoG)[2]、Edge Histogram (EH)[7]、Local Binary Pattern (LBP)[8]等，這些特徵在描述力、抗噪性與計算效率上各有其優缺點，而在辨識演算法部分，則多採用機器學習方式，如 Support Vector Machine(SVM)[9]、AdaBoost、K-Nearest Neighbor(KNN)、Local Sensitivity Hashing (LSH)[10]等，透過上述兩項組成元件的組合搭配與參數調校，開發出各式型態的物件辨識引擎。

   本研究的目標在於提出有效的物件辨識機制，而主要的作法則是尋找可靠的圖像特徵描述方式，

具體而言，我們將改善近來常被選用的區域二元化圖型(local binary patterns, LBP)，提出延展式的區域三元化圖型樣式(extended local ternary patterns, ELTP)，並針對其抗噪性、描述力與計算效率與 LBP 進行比較。除了描述力與抗噪性，運算效能也是必須納入考量的重要因素，LBP 之所以被廣為採用，有相當原因是因為其計算相當簡易，如要延伸原始 LBP 的功能，定義新的特徵描述方式，也要同時評估其計算的複雜度，以免顧此失彼。

本研究於第二章敘述相關之研究及其問題，第三章定義延展式區域三元化圖型及其中的 uniform patterns，第四章為抗噪力與描述力之實驗結果，第五章為結論與未來規劃。

2. 相關研究

2.1. 區域二元化圖型

區域二元化圖型(local binary patterns, LBP)是由芬蘭的 Oulu 大學 Ojala 等研究學者所提出的[8]，是一種區域紋理變化的特徵描述方式。原始的 LBP 演算法是在一塊 3x3 的區塊內，藉由中心點與參考點像素值的比較，將區塊內像素轉化為二元化表示後，再轉為十進位值。LBP 的計算流程如 Figure 1 所示，首先將 3x3 區塊中心點的像素值設為閾值，當周圍八個參考點的像素值小於閾值時，則將該點的值設為 0，反之則設為 1，之後再將套用閾值後的結果乘上對應的權重，即可算出區塊中心點的 LBP 值，而要描述一個影像區域時，則對區域內所有點進行 LBP 運算，再統計累加出影像的區域二元特徵直方圖(LBP Histogram)，作為圖像的特徵描述子。
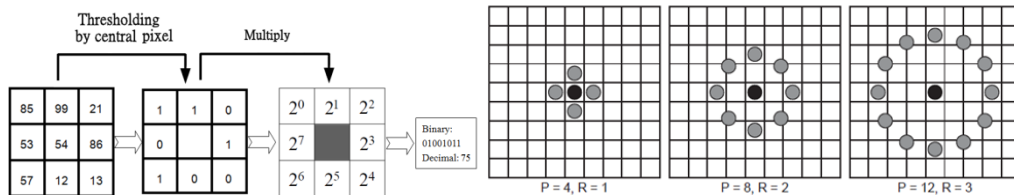
      (a)   (b)

Figure 1 (a)基本的 LBP 定義與運算方式；(b)可變化的 LBP(P, R)，其中區域半徑 R 與參考點數 P

上面所介紹的原始區域二元特徵所參考的區域只是一個 3x3 的區塊，支援的空間範圍較小，沒有辦法描述出較大的紋理結構，因此 Ojala 等學者把 LBP 的定義延伸，在延伸定義的 LBP 中，對於參考的區域大小以及參考點數不再局限於 3x3 的區塊中，使用者可依需求來改變參考的區域半徑 R 以及參考點數的點數(Figure 1(b))。假設中心點的座標為(0,0)，則參考點的座標為，R 為區域半徑，P 為參考點數。

LBP 最早被應用於材質分析(texture analysis)，之後陸續在人臉辨識、表情識別、多媒體搜尋、背景建模、動作分析等方面都有相當不錯的效果，由於 LBP 兼具效率與準確性的特點，也因此後續有若干的延伸或改善的版本，例如可處理不同角度的 LBP (rotation-invariant LBP)[8]、多重解析度的 LBP(multi-scale LBP) [11]、內含機率資訊的 Bayesian LBP[12]等。

在觀察 LBP 的計算過程中，可以發現幾個問題，第一是關於對雜訊的敏感度，尤其在平滑影像(例如：天空或海洋)的效果不理想，這是因為 LBP 選擇之閾值策略與平滑影像上的灰階值過於接近的緣故。假設 Figure 2(a)中間偏左的像素值受到了雜訊的干擾而有了微幅的變化(Figure 2(b))，依照 LBP 的定義，得到的二元樣式為 11001011，和原樣式的 Hamming distance 為 1，但化為十進位時卻有相當大的差距(75 vs. 203)，這也代表著在直方圖進行比對時，兩者會被視為不同的樣式群。
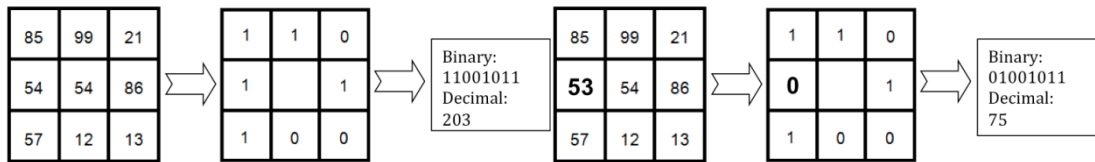
Figure 2 LBP 的抗噪力試驗說明

第二個問題則是樣式的個數與分群的方式，以最簡單的 LBP(8,1)來說，其區域二元特徵直方圖會有 256 個 bin，而 LBP(16,2)的區域二元特徵直方圖則遽增為 65536 個 bin，若處理的區塊大小為 32x32，在直方圖上得到的通常是極為稀疏的結果，是否具描述力與代表性值得探討，雖然類似的問題可以透過 uniform pattern 的概念稍微紓緩，然而當樣式的種類過多時，仍有樣式過多而樣本不足的根本問題。

2.2. 區域三元化圖型

針對 LBP 容易受到雜訊干擾這個問題，有研究學者提出三元化表示方式(Local Ternary Pattern，LTP)[13]，LTP 的主要作法，是將目前像素值與中心點像素值的關係，分為略大於(設為 1)、略小於(設為-1)和接近(設為 0)等三種狀況，亦即

$$LTP(i) = \begin{cases} 1 & p(i) - p(0) > \theta \\ 0 & if \quad |p(i) - p(0)| \leq \theta \\ -1 & p(i) - p(0) < -\theta \end{cases}$$

Equation 1 LTP 編碼方式

其中 P(i)為目前像素值，P(0)為中心點像素值，則為選定的門檻值，所形成的樣式由於涵蓋(-1, 0, 1)三種符號，所以被命名為區域三元化圖型。

以 Figure 2 為例，若接近的範圍定為 5，則可得到如 Figure 3(a)的表示方式，值得注意的是若在區塊中加上若干雜訊，並不會影響 LTP 的結果，所以 LTP 在某種程度上處理了雜訊干擾的問題。
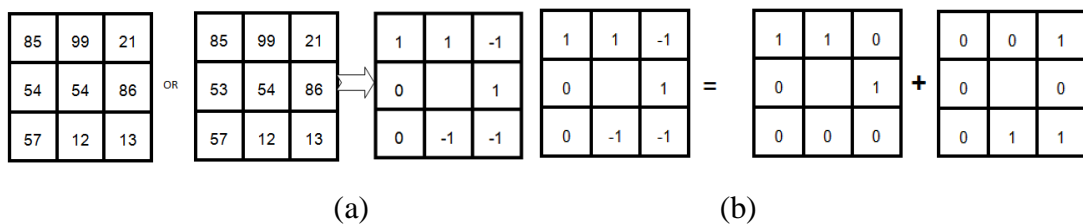

(a)   (b)

Figure 3 (a)Local ternary pattern 樣式描述方法(=5) (b)將 Local ternary pattern 拆解為兩組 LBP

然而這樣的編碼方式會造成樣式類別劇增；原本 LBP(8,1)有 256 種組合，而 LTP (8,1)則有 6561 種，LTP (16,2)的樣式更高達 43,046,721 種，前述 LBP 直方圖稀疏的問題，在這種編碼方法下只會更為嚴重。而為了避免這個問題，原始的 LTP 在編碼完後會拆解為兩組 LBP，其中一組保留+1 的樣式，另一組則保留-1 的樣式，如此一來每組的樣式種類就只有 256 種，至於後續的圖像相似度計算，也採取分開比對的方式，然後再以平均值代表。(Figure 3(b))

LTP 雖然加強了圖型的抗噪力，但同時也產生了若干懸而未決的問題，首先是門檻值的選取方式，固定的門檻值並不是一個好方法，在不同的情況下，適合的門檻值也不盡相同，要是改變使用環境，可能會使效果大打折扣，若是使用平滑區域(texture-free region)的統計資訊或可變式的(adaptive)門檻設定，效果應當較佳。再來是編碼方式，使用三元的描述法的目的在於抗噪與強化特徵描述能力，但是

在 LTP 拆解的過程，則又有把原屬不同樣式的合為同一群組的效果，一來一往之間的轉換是否合理，令人質疑，而根據我們初步實驗的結果，在某些情況下，使用此法定義的 LTP 表現有時尚不如原始的 LBP，因此如何將 LTP 作有意義的降維群組化，也是本研究的重要目標。最後是直方圖稀疏的問題，雖 LTP 在編碼完後會拆解成兩組的 LBP，但若是增加參考點的數量，所拆解出的 LBP 還是存在著直方圖稀疏的問題，直方圖過於稀疏的問題在 LTP 中依然沒有得到解決。也因此我們將根據上述所提的特徵描述子的缺點，提出改進的特徵描述子：延展式區域三元化圖型。

3. 延展式區域三元化圖型

本研究的目標是改善區域二元化圖型的抗噪性與描述性，期望應用於物件分類或辨識時能有較佳的效能與準確性。我們將新提出的描述子並命名為延展式區域三元化圖型(Extended Local Ternary Patterns, ELTP)，在定義 ELTP 時，有幾個問題必須注意，包含三元化的標準、ternary pattern 對應至十進位數字的方式、ternary pattern 降維的可能作法、不同 ELTP 間相似度的計算方法等。此外，轉換為三元表示法後導致樣式種類大量增加(LBP(8,1)有 256 個 patterns，LTP(8,1)則有 6561 個 patterns)，如何進行有系統的降維，則是 ELTP 成敗的關鍵，以下我們將依序針對前述問題討論可能的作法。

3.1. 三元化的範圍設定

在 LTP 的定義中，將中心點與鄰近點像素值的關係分為三類的作法其實相當直覺，也就是分成大於、小於及接近等三種範圍，最簡單的作法是設定固定範圍，也就是在公式(3.1)中取固定的θ值

但是關於門檻值θ的選擇，其實需要更進一步來討論。就抗雜訊的角度而言，θ值應和雜訊的強度有關，所以應由統計結果來決定θ的值(例如$\alpha * \sigma$)。我們採用的方法是根據 local statistics 來設定門檻，先在每一個處理的參考區塊內計算出區塊中心點與參考點的標準差，再對標準差取一定的比例當作門檻值(見 Equation 2)。根據此方法，我們將 ELTP 的編碼方式定義如下：為了決定門檻值，我們先算出區塊內中心點與參考點的標準差，再將所算出的標準差乘以σ，算出的值即為門檻值，利用此門檻值即可得到一組三元碼(ternary code)。因此以 Equation 2 ELTP 編碼方式   Figure 4 為例，門檻值為 25.94×0.3=7.7 (α設為 0.3)。

$$LTP(i) = \begin{cases} 2 & p(i) - p(0) > \theta \\ 1 & if \quad |p(i) - p(0)| \le \theta \\ 0 & p(i) - p(0) < -\theta \end{cases} \quad \theta = \alpha * \sigma$$

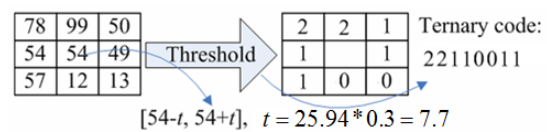Equation 2 ELTP 編碼方式   Figure 4 ELTP 門檻值設定

3.2. 樣式編碼與轉換方式

在將區域內容轉換為三元表示法後，得到的是一組三元碼，若直接用這組三元碼來統計累加直方圖，會產生直方圖過於稀疏的問題。以 ELTP(8,1)為例，所得到的三元樣式會高達 6561 種，為了避免這種問題，我們把直方圖內的 bins 做合併來降低直方圖的維度。事實上，存在著數種降低直方圖維度的方式，包括直接將相鄰 bins 合併，但缺乏考慮相鄰的 bin 的 hamming distance 關係；其二是將三元碼換底對應至二進制碼，好處是維度下降了，但無法確保每個 Bins 合併三元碼的數量，且只能降維至2^n。

最後一種方式則是利用 bin 跟 bin 的關係來做合併，也就是把關係密切的 bin 視為同一群，以達到降維的效果。由於每一個 bin 代表一個三元碼，對 bin 作分群就等同於對三元碼作分群，因此我們把每

一個三元碼視為一個 pattern，依據 pattern 與 pattern 間的相似度來做分群，而分群的數目就是我們要降的維度。使用這種方式來做降維，對我們所要降的維度不會有所限制 ，而所分出來的群之間有著不同的大小，也因為是利用 pattern 間的相似度來合併 bins 而有了合理的解釋。

由於資料表示方法的特殊性，資料分群中常用的 k-means 或 mean-shift 等演算法並不適用於字串型態的輸入樣本。一般針對的字串型態的資料，由於缺乏幾何空間的對應，因此多採取 graph representation 的作法(Figure 5)，將每一字串視為 graph 上的節點，而節點與節點間的連結強度，則是字串間的相似度。因此我們使用只需相似矩陣就可進行分群的 spectral clustering 相關演算法進行分群。
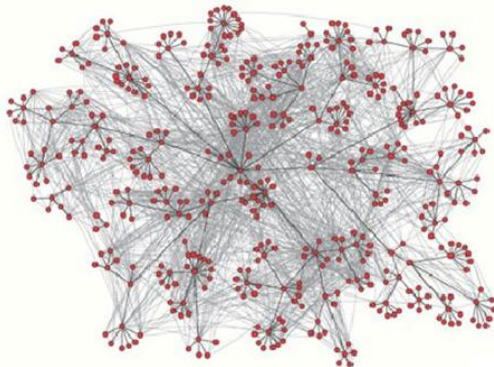

Figure 5 以 graph 方式表示樣式間的關係[14]

在算 ELTP 的三元編碼時，根據 local statistics 來設定門檻，可避免像素受到雜訊干擾而影響編碼的結果，若是所得的三元編碼受到雜訊的影響而改變了編碼的結果，也可經由 spectral clustering 的分群降維，把受到干擾與未受到干擾的三元編碼分為同一群，使得在統計特徵向量直方圖時不會受到影響。

3.3. Spectral Clustering

Spectral Clustering 是用圖論(Graph Theory)的角度來看分群，它把要分群的資料視為一個有權重的無向圖(Undirected weight graph)，一個節點代表一筆資料，邊上的權重代表資料間的相似度，分群就是要把圖形切成數個不相交的子圖(Figure 6(a))，讓同一子圖內有著較高的權重值，而不同子圖間的權重值則較低。
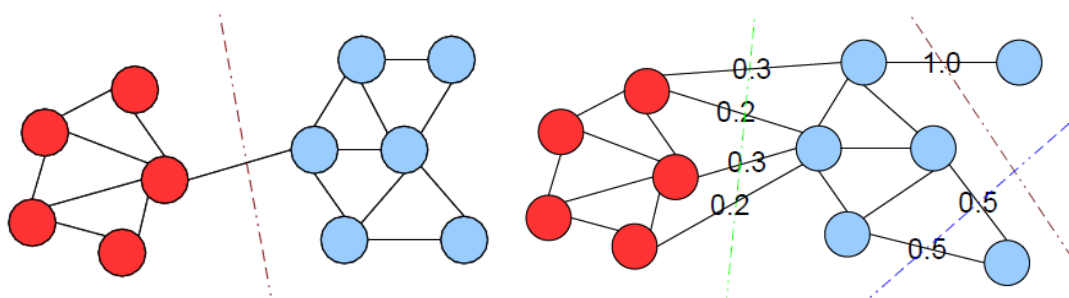

Figure 6 (a)圖形切割式意圖;(b)圖型切割與孤立點

在圖形切割(Graph Cut)的問題中，我們要將圖形的一些邊給切斷，讓圖形被分割成數個獨立聯通的子圖，對於那些被切斷的邊的權重總和我們稱之為 Cut 值。

$$A, B \in V, A \cap B = \emptyset$$

$$cut(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

Equation 3 Cut 值

8

而圖形切割的問題就是將圖形切成數個子圖，並要求切割所得的 Cut 最小。對於圖形切割的問題，若是只考慮到 Cut 值，而忽略了切割區域的大小，極有可能發生 Figure 6(b)的情形，也就是孤立點的問題，然而為了讓切割區域大小相對均勻，可以使用的目標函數(Objective function)有 RatioCut 與 NormalizeCut。雖然上述兩種函數可避免孤立點的問題，但是要最小化它們卻是屬於 NP-hard 的問題。我們可以藉由放寬(Relax)指示向量(Indicator Vector)，使得對圖型切割的目標函數求極值的問題轉化成了 Rayleigh Quotient 求極值的問題。

　　根據上述所介紹的特性與概念，我們可以了解到 Spectral Clustering 有哪些重要的步驟(Figure 7)，根據這些步驟，我們可以得到 Spectral Clustering 演算法的雛型。
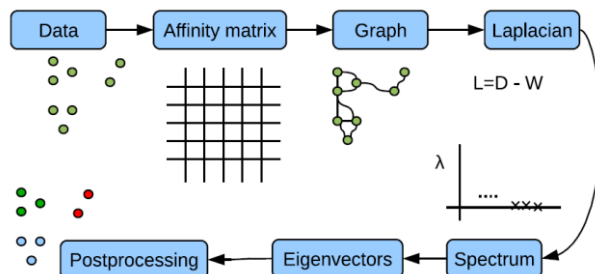

Figure 7 Spectral Clustering 流程圖

　　下面是三個較常被使用的 Spectral Clustering 演算法，分別是 Un-normalized Spectral Clustering(Figure 8)、Normalized Spectral Clustering(Figure 9)與 NJW algorithm(Figure 10)[15]。

Unnormalized spectral clustering

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.
- Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L.
- **Compute the first k eigenvectors $u_1, \ldots, u_k$ of L.**
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
- Cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j | y_j \in C_i\}$.

Figure 8 Un-normalized Spectral Clustering 演算法

Normalized spectral clustering according to Shi and Malik (2000)

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.
- Construct a similarity graph by one of the ways described in Section 2. Let W be its weighted adjacency matrix.
- Compute the unnormalized Laplacian L.
- **Compute the first k eigenvectors $u_1, \ldots, u_k$ of the generalized eigenproblem $Lu = \lambda Du$.**
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
- For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
- Cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j | y_j \in C_i\}$.

Figure 9 Normalized Spectral Clustering 演算法

```
Normalized spectral clustering according to Ng, Jordan, and Weiss (2002)
Input:   Similarity matrix S ∈ ℝⁿˣⁿ, number k of clusters to construct.
● Construct a similarity graph by one of the ways described in Section 2.
  Let W be its weighted adjacency matrix.
● Compute the normalized Laplacian L_sym.
● Compute the first k eigenvectors u₁,...,uₖ of L_sym.
● Let U ∈ ℝⁿˣᵏ be the matrix containing the vectors u₁,...,uₖ as columns.
● Form the matrix T ∈ ℝⁿˣᵏ from U by normalizing the rows to norm 1,
  that is set t_ij = u_ij/(∑ₖ u²_ik)^(1/2).
● For i = 1,...,n, let y_i ∈ ℝᵏ be the vector corresponding to the i-th row of T.
● Cluster the points (y_i)_{i=1,...,n} with the k-means algorithm into clusters
  C₁,...,Cₖ.
Output:  Clusters A₁,...,Aₖ with A_i = {j| y_j ∈ C_i}.
```

Figure 10 NJW 演算法

Table 1 分群結果

	g1-128	g1-256	g2-128	g2-256	h2-128	h2-256	h2.5-128	h2.5-256
Mean Distance	3.134	2.615	3.17	2.667	3.459	2.94	3.389	2.65
Max Cluster Size	68	38	70	37	73	38	69	48
Min Cluster Size	39	14	33	15	38	14	36	17
Standard Deviation	5.66	4.41	7.16	4.28	6.45	4.6	6.04	4.53

Table 1 為上面分群結果的數據，g 代表使用高斯模型定義相似度的分群結果，s 代表使用我們所定義的相似度所得的分群結果，平均距離是依每一群內的平均距離來計算，代表著分群結果的和諧度，而標準差是依據每一群內 pattern 的個數來計算，代表著分群結果的均衡度。未分群前 pattern 間的平均距離為 7.11，分成 256 群時平均每群內有 25.62 個 pattern，分成 128 群時則有 51.26 個 pattern。

從表中可觀察到，不管使用何種相似度定義來做分群，所得的分群結果並沒有過大的差異。在參數設定上，則可以觀察到使用的參數若越能加大 pattern 間相似度的差異，則所得到的分群和諧度也會越高。

Spectral Clustering 計算複雜度約為 Eigensystem 加上 k-means 的計算複雜度，若迭代次數為 1，則複雜度約為 O(dcN)+O(NK)，c 為 dxN 的資料矩陣中平均每行非零項數，K 為所求的分群數。由於計算繁瑣，因此所有的分群結果皆事先完成並存起來，在計算 ELTP 時再讀取分群結果。

4. ELTP 中的 Uniform Patterns

Uniform Pattern 是 LBP 中一個重要的性質，更是廣泛的使用在許多 LBP 的應用上。不過由於 ELTP 是使用三元編碼，pattern 數比二元編碼的 LBP 多上許多，對於 LBP 中所定義的 uniform pattern，我們應該如何去定義屬於 ELTP 的 uniform pattern，本章節中將探討這個重要議題。

4.1. LBP 中的 Uniform Pattern

Uniform pattern 的概念是 Ojala 等學者於 2002 年所提出的[8]，他們發現在 LBP 的 pattern 中，有著大量且型式特定的 pattern 存在於紋理的圖像中，這些 pattern 之所以被稱為 uniform pattern，是因為它們有著共通的一點：在環狀的二元編碼中，最多只有 2 個 0 到 1 或 1 到 0 的轉換，例如：0000011、11001111 和 11111111 等都屬於 uniform pattern。

在經過紋理圖像實驗後，Ojala 等人發現若使用 LBP(8,1)，uniform pattern 最高可佔有 90%，若是使用 LBP(16,2)，最高則有 70%。在 LBP(8,1) 中，全部共有 256 種 pattern，其中 58 種 pattern 是屬於 uniform pattern，佔 23%。若使用 uniform pattern 當作特徵向量，則統計累加 58 種 uniform pattern 與其它非 uniform pattern 的直方圖作為特徵向量，而特徵向量的維度為 59 維，與未使用 uniform pattern 的 256 維相比之下，維度要小得多。

4.2. ELTP 中的 Uniform Pattern

在 ELTP(8,1) 中總共有 6561 種 pattern，我們要從這些 pattern 中找出部分特定的 pattern 其大量出現於影像中。我們首先用類似 LBP 中 uniform pattern 的定義來定義 ELTP 中的 uniform pattern。我們將使用紋理與一般的圖像作紋理圖像實驗(Figure 11)，檢查圖像中 uniform pattern 所佔的比例，以驗證我們所定義的 uniform pattern 是否大量的存在於圖像中。
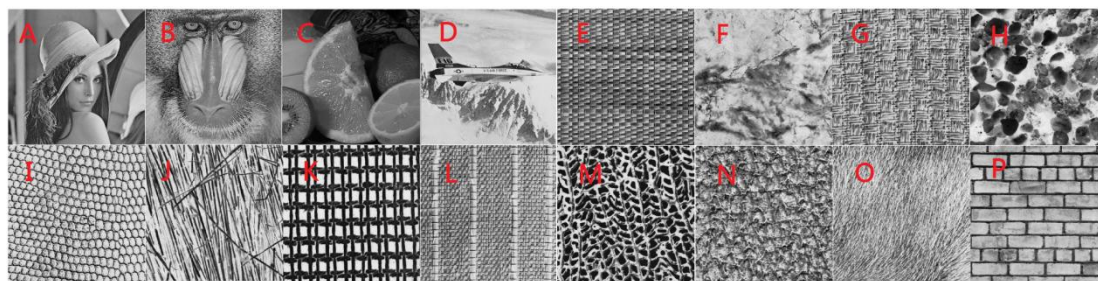
Figure 11 紋理圖像實驗圖像

定義一(UELTP1)：在環狀的三元編碼中，相鄰 digit 最多只有 2 次的變換(transition)。例如：00022000、21122222、12111111(Figure 12(a))。在 6561 個 pattern 中有 171 個 pattern 屬於定義一的 uniform pattern。

定義二(UELTP2)：在環狀的三元編碼中，相鄰兩個 digit 間 Hamming distance 總和不得超過 2(Figure 12(b))。例如：11112111、21112222、12111111。在 6561 個 pattern 中有 115 個 pattern 屬於定義二的 uniform pattern。

定義三(UELTP3)：在環狀的三元編碼中，最多只有 4 次 digit 的變換(transition)(Figure 12(c))。例如：22001222、22212112。

定義四(UELTP4)：在環狀的三元編碼中，相鄰兩個 digit 間 Hamming distance 總和不得超過 4(Figure 12(d))。例如：20011112、22100012。值得注意的是，藉由定義四所得之 uniform pattern 是定義三所得 uniform pattern 之子集合。

Figure 12unIform pattern (a)定義一 (b)定義二 (c)定義三 (d)定義四 範例

Table 2 紋理圖像實驗結果 LBP 與 ELTP 中 uniform pattern 百分比

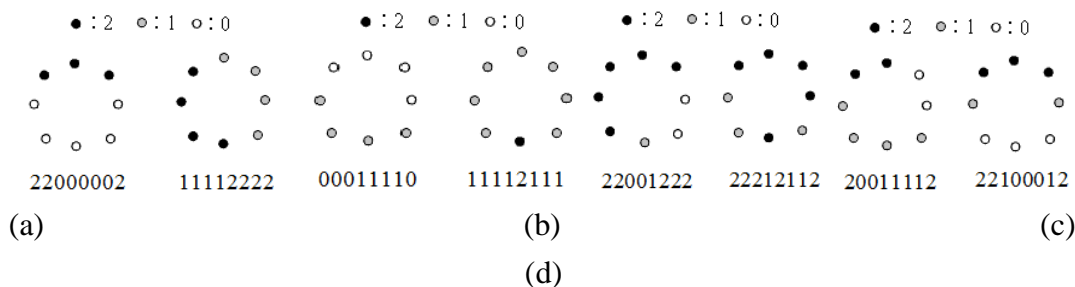|	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
ULBP	0.78	0.68	0.85	0.82	0.83	0.75	0.82	0.81
UELTP1	0.25	0.26	0.23	0.28	0.33	0.28	0.33	0.28
UELTP2	0.21	0.22	0.18	0.23	0.29	0.24	0.3	0.25
UELTP3	0.79	0.71	0.86	0.83	0.87	0.77	0.88	0.83
UELTP4	0.72	0.62	0.81	0.78	0.82	0.72	0.84	0.79
	I	**J**	**K**	**L**	**M**	**N**	**O**	**P**
ULBP	0.76	0.8	0.85	0.82	0.84	0.77	0.67	0.76
UELTP1	0.39	0.39	0.45	0.34	0.48	0.26	0.25	0.31
UELTP2	0.35	0.35	0.43	0.31	0.45	0.23	0.21	0.27
UELTP3	0.84	0.9	0.89	0.87	0.9	0.78	0.74	0.8
UELTP4	0.8	0.86	0.86	0.83	0.88	0.72	0.66	0.73

　　由 Table 2 我們可以觀察到，我們所定義的 ELTP 的 uniform pattern 在圖像中並沒有大量的出現，最多也只有佔到 48%，與 LBP 的 uniform pattern 相較之下少了許多。另外定義三與定義似的 uniform pattern 都大量出現於圖像中，其中最高可佔到 90%。在所有的 6561 種 pattern 中，有 1767 種 pattern 是屬於定義三的 uniform pattern，佔 27%，而屬於定義四的 uniform pattern 則有 1067 種，佔 16%。相較之下，定義三多了 700 個 pattern，但出現的百分比僅相差 4%-11%，因此依定義四所得的 uniform patterns 應屬較有效率的表示方法。

　　相較於 LBP 中的 uniform pattern 個數 58 個，前面所定義的兩種 uniform pattern 數量都高達一千種以上，若是直接使用這些 pattern 的直方圖做為特徵向量，還是可能會有直方圖稀疏的問題，因此我們將在下一節討論直方圖降維的可能做法。

4.3. Uniform Pattern 的降維

　　把 uniform pattern 作分群來降維，它的概念與第三章所用到的 pattern 分群降維法並無不同，都是每個 bin 代表著一個 pattern。而 pattern 間的距離，我們採用 hamming distance 來定義，在算出 pattern 間的距離後，我們會利用距離來計算 pattern 間的相似度，相似度的定義如下：

$$S = e^{-\frac{D^2(x,y)}{2\sigma^2}} \qquad \text{OR} \qquad S = (1 - \frac{D(x,y)}{2n})^{\gamma}, \gamma > 1$$

Equation 4 Hamming Distance 相似度定義

　　在算得 pattern 的相似矩陣(similarity matrix)後，我們使用 spectral clustering 來進行分群。對於那些不屬於 uniform pattern 的 pattern，我們認為它們還是能提供資訊，因此我們把它們都視為於同一類的 pattern，但是它們並不參與 pattern 的降維，因此特徵向量的維度會為：uniform pattern 所降至的維度＋1。

5. 實驗結果

在第三章與第四章時，我們介紹了我們所改進的特徵描述子，在此章節中，我們設計了三種實驗來測試特徵描述子的抗噪力。第一個實驗設計為取一圖像中的若干區塊，比較其加入雜訊後直方圖的變化。實驗二為取一圖像中的區塊，比較該區塊受到光影干擾後直方圖的變化。實驗三為取一圖像中的區塊，在區塊中將入雜訊，且逐漸增加雜訊的強度，最後與未受到雜訊干擾的直方圖做比較。
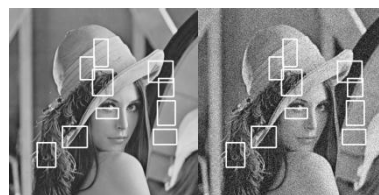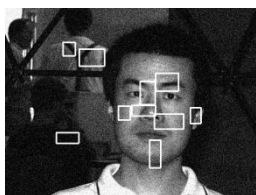
5.1. 抗噪力實驗(一):加入高斯雜訊

抗噪力實驗(一)的實驗設計為先隨機選取一張影像的 10 個區塊，計算那些區塊特徵描述子的直方圖，然後對影像加上雜訊，再計算相同區塊特徵描述子的直方圖，觀察原圖與加入雜訊後的直方圖變化，並用 Chi-Square distance 與直方圖相交(Histogram Intersection)來計算兩組直方圖之間的相似度，最後把每個區塊的相似度加總平均，即可判斷抗噪力的強弱。

在第一個抗噪力實驗中,參與實驗的特徵描述子皆為(8,1)的設定,即參考半徑為1與參考點數為8,加入雜訊的種類為高斯雜訊,雜訊強度約為 SNR 值 15,而參與實驗的特徵描述子分別為:

LBP、ULTP、UniformLBP、DLTP)與做分群的降維的 ELTP(Gaussian 表示用高斯模型定義相似度、Hamming 表示用我們自行定義的相似度,數字部分代表所降至的維度);與第三種定義的 ELTP uniform pattern(UELTP3,G 表示用高斯模型定義相似度,H 表示用我們自行定義的相似度)。 與第四種定義的 ELTP uniform pattern(UELTP4,G 表示用高斯模型定義相似度,H 表示用我們自行定義的相似度)。
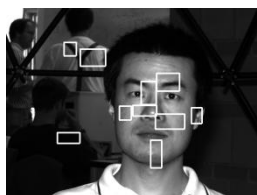
實驗影像一   實驗影像二

抗噪力實驗(一):加入高斯雜訊實驗結果

描述子\相似度	實驗影像一		實驗影像二	
	Chi-Square	**直方圖相交**	**Chi-Square**	**直方圖相交**
LBP	0.451	0.585	0.391	0.604
ULTP	0.853	0.388	0.579	0.497
UniformLBP	0.418	0.594	0.336	0.649
DLTP	0.701	0.452	0.582	0.499
Gaussian64	*0.238*	0.702	*0.270*	0.693
Gaussian128	0.312	0.675	0.305	0.675
Gaussian256	0.368	0.633	0.351	0.628
Hamming64	0.258	0.718	0.255	0.705
Hamming128	0.245	0.713	0.244	*0.714*
Hamming256	0.328	0.659	0.333	0.641
UELTP3-G58	0.361	0.63	0.319	0.659
UELTP3-G128	0.387	0.616	0.362	0.626
UELTP3-G256	0.412	0.599	0.366	0.603
UELTP3-H58	0.341	0.639	0.290	0.672
UELTP3-H128	0.362	0.637	0.342	0.638

UELTP3-H256	0.397	0.607	0.353	0.607
UELTP4-G58	0.378	0.628	0.321	0.659
UELTP4G128	0.4	0.605	0.327	0.634
UELTP4-G256	0.426	0.59	0.315	0.613
UELTP4-H58	0.391	0.612	0.325	0.651
UELTP4-H128	0.395	0.609	0.331	0.634
UELTP4-H256	0.433	0.588	0.341	0.602

在抗噪力實驗(一)的實驗結果可觀察到，與 LBP 和 LTP 相比，ELTP 有較佳的抗噪能力。在欲降至的維度選擇，降至 256 維時抗噪力並沒有特別突出，但其餘的維度(58,128)皆有不錯的抗噪力。

5.2. 抗噪力實驗(二):光影變化

在抗噪力測試(二)的實驗設計為取兩張相同的影像，其中一張有受到光影的干擾，手動框選受到干擾的區塊並計算其特徵向量直方圖，並使用 Chi-Square distance 與直方圖相交來計算受干擾前與干擾後的差異。參與實驗的特徵描述子與抗噪力測試(一)相同。

實驗影像三   實驗影像四

抗噪力實驗(二):光影變化實驗結果

描述子\相似度	實驗影像三		實驗影像四	
	Chi-Square	直方圖相交	Chi-Square	直方圖相交
LBP	0.189	0.732	0.318	0.631
ULTP	0.139	0.773	0.553	0.476
UniformLBP	0.141	0.761	0.296	0.678
DLTP	0.13	0.777	0.582	0.458
Gaussian64	*0.083*	*0.831*	*0.172*	*0.767*
Gaussian128	0.099	0.828	0.225	0.719
Gaussian256	0.16	0.767	0.237	0.685
Hamming64	0.084	0.829	0.154	0.766
Hamming128	0.106	0.812	0.182	0.747
Hamming256	0.158	0.771	0.218	0.698
UELTP3-G58	0.129	0.787	0.202	0.731
UELTP3-G128	0.156	0.754	0.228	0.707
UELTP3-G256	0.182	0.731	0.242	0.683
UELTP3-H58	0.111	0.782	0.206	0.72
UELTP3-H128	0.155	0.76	0.223	0.718
UELTP3-H256	0.179	0.74	0.236	0.673
UELTP4-G58	0.134	0.773	0.195	0.736

UELTP4G128	0.167	0.751	0.221	0.721
UELTP4-G256	0.182	0.733	0.252	0.682
UELTP4-H58	0.138	0.763	0.186	0.75
UELTP4-H128	0.138	0.761	0.241	0.698
UELTP4-H256	0.157	0.732	0.252	0.678

在抗噪力實驗(二)的結果中，對於光影變化的影響，ELTP 均有著較佳的抵抗能力，在所降至的維度方面，降至 256 維對光影的抵抗力較其它維度差，相對而言不是一個理想的維度。對於這三組特徵描述子的比較結果， 不論是一般的 ELTP 與 ELTP 的 uniform pattern 對雜訊的抵抗力相當的好，相較於 LBP 的 uniform pattern 或是一般的 LBP 都是略勝一籌。

6. 結論及未來工作

本研究的目標是改善區域二元化圖型的抗噪性與描述性，期望應用於物件分類或辨識時能有較佳的效能與準確性。我們將新提出的描述子並命名為延展式區域三元化圖型。在定義 ELTP 時，對於轉換為三元編碼後導致樣式種類大量增的問題，我們利用了 Spectral clustering 有效的解決了這個問題，並且成功的提升了 ELTP 的抗噪能力。對於 LBP 中 Uniform pattern 的存在，我們也在 ELTP 中找出三元編碼的 Uniform pattern，並對三元編碼的 Uniform pattern 數量過多的問題，有效的使用了 Spectral clustering 來解決。在第五章的實驗中，我們驗證了 ELTP 的抗噪能力，而在第六章的圖型識別應用中，不論是材質分析或是人臉辨識，ELTP 都出色的結果。

對於 UELTP1 的 Uniform pattern 來說，其 pattern 數只佔了 2.5%，但是卻可在圖像中佔到 20~40%。對於這種特性若使用分群將其聚合，似乎有些可惜，因此未來在對於三元編碼的 Uniform pattern 的定義與降維，可對 UELTP1 做更進一步的研究，看其能否更進一步的增強抗噪力與描述力。

此外，本研究所提出的 ELTP 乃基於原始 LBP 的定義，但是 LBP 本身也有許多的變形，如 Rotation-invariant LBP[8]、XY-LBP[19]、volume LBP (VLBP) [20]、CS-LBP[21]等，如何將本論文所提出三元編碼的概念套用至上述的描述子中，將是值得後續研究探討的問題。

7. 參考文獻

[1] P. Viola and M. Jones, "Robust Real-Time Object Detection". Proc. ICCV Second Int'l Workshop Statistical and Computational Theories of Vision Modeling, Learning, Computing, and Sampling, July 2001.

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '05), vol. 1, pp. 886-893, 2005.

[3] R.Plamondon and S.N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey". IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63-84, Jan. 2000

[4] Google goggles, www.google.com/mobile/goggles

[5] VOC 2009 Challenge Results:, http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2009/results/index.html

[6] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int'l J. Computer Vision, vol. 2, no. 60, pp. 91-110, 2004.

[7] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada ,"Color and texture descriptors," IEEE Trans. Circuit Syst. Video Technol., vol. 11, pp. 703–715, June 2001

[8] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, July 2002.

[9] Pontil and A. Verri, "Support Vector Machines for 3D Object Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20(6), pp. 637-646, 1998.

[10] A. Gionis, P. Indyk, and R. Motwani, "Similarity Search in High Dimensions via Hashing," Proc. Very Large Data Base Conf. (VLDB '99), pp. 518–529, Sept. 1999.

[11] T. Maenpaa, and M. Pictikainen, "Multi-scale binary patterns for texture analysis," Springer Berlin / Heidelberg, 2003.

[12] C. He, T. Ahonen and M. Pietikäinen, "A Bayesian Local Binary Pattern texture descriptor",Proc. Int l Conf. on Pattern Recognition, 2008.

[13] X. Tan and B. Triggs. "Enhanced local texture feature sets for face recognition under difficult lighting conditions". In Analysis and Modeling of Faces and Gestures, volume 4778 of LNCS, pages 168–182. Springer, 2007

[14] Matthias Hein and Ulrike von Luxburg ,"Short Introduction to Spectral Clustering", MLSS 2007

[15] Ng, A., Jordan, M., and Weiss, Y. (2002). On spectral clustering: analysis and an algorithm. In T. Dietterich,S. Becker, and Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems 14 (pp. 849 –856). MIT Press.

[16] Brodatz database, http://www.ux.uis.no/~tranden/

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp.2037-2041, Dec. 2006.

[18] The Yale Face Database B , http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html

[19] G. Zhao and M. Pietik¨ainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. PAMI, 29(6):915–928, 2007.

[20] G.Zhao and M. Pietikäinen, "Dynamic Texture Recognition Using Volume Local Binary Patterns", Proc. ECCV 2006 Workshop on Dynamical Vision, Graz, Austria, 2006.

[21] M. Heikkil¨a, M. Pietik¨ainen, and C. Schmid, "Description of interest regions with center-symmetric local binary patterns", In Computer Vision, Graphics and Image Processing, 5th Indian Conference, pages 58–69, 2006.

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

---

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

   ■達成目標

   □ 未達成目標（請說明，以 100 字為限）

       □ 實驗失敗

       □ 因故實驗中斷

       □ 其他原因

   說明：Extended local ternary patterns (ELTP)作為區域描述之特徵，兼具描述力、抗噪性等優點，在計算複雜度部分與 local binary patterns 差異不大，透過材質辨識、人臉辨識等實驗，驗證了 ELTP 在物件偵測能力之優越性。

2. 研究成果在學術期刊發表或申請專利等情形：

   論文：■已發表 □未發表之文稿 □撰寫中 □無

   專利：□已獲得 □申請中 □無

   技轉：□已技轉 □洽談中 □無

   其他：（以 100 字為限）

   共有兩篇論文發表於國際會議:

   1. W. Liao, "Region Description Using Extended Local Ternary Patterns", Proceedings of the 20th International Conference on Pattern Recognition, pp. 1003-1006, 2010.

   2. W. Liao and T. J. Young, "Texture Classification Using Uniform Extended Local Ternary Patterns", Proceedings of The IEEE International Symposium on Multimedia, pp. 191-195, 2010.

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以500字為限）

本研究所提出之 extended local ternary pattern，乃基於著名的區域描述法 local binary pattern，期望在不增加太多計算量的前提之下，改善原有方法之抗噪力，使其適用於各式之光源條件與運作環境。

三元化的想法，並非本研究肇始，然而先前提出的定義方式，只有點出問題，而沒有真正的解決問題，關鍵的降維概念並未被導入，以致於只能紙上談兵，甚或使用了錯誤的降維方式，反而使得整體的效能其準確度都降低。

本研究使用 spectral clustering 的分群方式，作為降維之手段，不但具有通用性，也有應用上之彈性，此外，本研究亦探討這類型描述子的重要的性質:uniform pattern，在新的定義基礎上，重新檢視 uniform pattern 的組成與分布，透過這個面向的研究，豐富了未來發展的可能性。

以學術貢獻而論，目前本研究之成果已發表於兩個國際知名會議：Internal Conference on Pattern Recognition（ICPR 2010），International Symposium on Multimedia（ISM 2010），提出之立論獲 reviewer 的高度贊同，後續的一些成果正在整理當中，將有更多的發表出現。

以技術創新的層次而言，ELTP 要解決的是電腦視覺中物件描述與辨識的問題，在實務上有許多重要的應用。從本研究的實驗數據顯示，不管是材質分類、人臉辨識，甚或背景建模等，ELTP 均有著優於 LBP 的表現，因此將可有效作為物件辨識核心的主要特徵描述方法。

就未來的發展而言，因 ELTP 乃基於 LBP，而 LBP 本身也有許多的變形，如 Rotation-invariant LBP、XY-LBP、volume LBP（VLBP）、CS-LBP 等，如何將本研究所提出三元編碼的概念套用至上述的描述子中，將是值得後續研究探討的問題。

國科會補助專題研究計畫項下出席國際學術會議心得報告

日期：100 年 10 月 30 日

計畫編號	NSC 99－2221－E－004－012－		
計畫名稱	基於延展式區域三元化樣式的物件描述方法		
出國人員 姓名	廖文宏	服務機構 及職稱	國立政治大學資訊科學系
會議時間	99 年 8 月 23 日至 99 年 8 月 26 日	會議地點	土耳其伊斯坦堡
會議名稱	(中文)第二十屆國際圖型識別研討會 (英文)20th International Conference on Pattern Recognition		
發表論文 題目	(中文)基於延展式區域三元化樣式的區域描述方法 (英文) Region Description Using Extended Local Ternary Patterns		

20th International Conference on Pattern Recognition (ICPR 2010)論文接受函

Message from The PaperCept Conference Management System

Message originated by Mujdat Cetin

Dear Prof. Wen-Hung Liao,

We are pleased to inform you that the paper

Submission number: 1214
Title Region Description Using Extended Local Ternary Patterns
Authors: Wen-Hung Liao*

you have submitted to ICPR 2010 has been accepted. The conference program will be announced soon and will contain information on whether your paper will be presented in an oral or a poster session.

The referee reports for your paper are included with this message. Please note that the referees have spent considerable time reviewing your paper. Therefore we strongly recommend that you take their comments into

account when preparing the camera-ready version of your manuscript.

Final manuscripts should be up to 4 pages long. However authors may purchase one extra page at 100 EUR if they find it necessary to address the concerns of the reviewers. You will receive a separate e-mail message on the procedure for submitting the final, camera-ready version of your manuscript. The deadline for final paper submission is May 15, 2010.

Please note that at least one of the authors must register for the conference paying regular registration fees (not the student rate) before submitting the camera-ready version of his/her manuscript(s). Papers submitted without registration will not be included in the conference program or in the proceedings. According to IAPR's policy, should an author have more than one paper accepted, only one registration is required for publication although other authors are encouraged to register and participate in the conference as well. The deadline for Early/Author
Registration is May 15, 2010.

Each paper (oral or poster) must be presented by an author at the conference. IAPR/IEEE reserves the right to exclude a paper from distribution after the conference (e.g., removal from IEEE Xplore) if the paper is not presented at the technical meeting. In particular, a paper accepted for the ICPR 2010 technical program will not be published on IEEE Xplore if its authors or a proxy do not present the paper at the conference. Please also see the IAPR Statement of Ethics and IAPR Ethical Requirements for Authors on
http://dev.conbrio.com/iapr/site/news/index.php#295 .

Please note the information below on IAPR Travel Stipends, and please visit the conference web site(http://www.icpr2010.org) for information on the Conference Program, Registration, and Accommodation.

We hope to see you at ICPR 2010 in Istanbul on August 23-26, 2010.

Sincerely,

Kim Boyer, Mujdat Cetin, Seong-Whan Lee
Technical Program Co-Chairs, ICPR 2010
icpr2010technicalprogram@lists.sabanciuniv.edu

ICPR 2010 心得報告

2010 年 8/23-26 日在土耳其伊斯坦堡會議與展覽中心(Istanbul Convention & Exhibition Centre)舉行的國際圖型識別研討會(International Conference on Pattern Recognition)，每兩年舉辦一次，至今已是第二十屆，堂堂邁入第四十年，是圖形識別領域歷史最為悠久，也是最重要的會議之一。

土耳其是昔日東西文化交會之處，首都伊斯坦堡有著豐富的歷史人文遺跡，也遍佈象徵回教信仰的清真寺，因此會議投稿的狀況相當踴躍，也許是為了趁參與會議之便，順便到此一遊。

本次會議分為六大主題 (tracks)，共有 2140 篇投稿，接受的比例在五成上下，實際與會的人數達一千多人，可以說規模盛大，而報告內容也是包羅萬象，吸引來自不同領域與國家的研究學者。此外，本次會議也舉辦了許多競賽，包含:

- Semantic Description of Human Activities
- Pattern Recognition in Histopathological Images
- Classifier Domains of Competence
- Graph Embedding for Pattern Recognition
- Bi-modal Handwritten Text Recognition
- The CLEF Cross Language Image Retrieval Track
- Verification of Video Source Camera Competition
- Mobile Biometry Face and Speaker Verification Evaluation

此次會議的六大主題為:

Track I	Computer Vision
Track II	Pattern Recognition and Machine Learning
Track III	Signal, Speech, Image and Video Processing
Track IV	Biometrics and Human Computer Interaction
Track V	Multimedia and Document Analysis, Processing and Retrieval
Track VI	Bioinformatics and Biomedical Applications

由於有六場論文發表同步進行，加上海報部份也在相鄰場地展示，因此實際參與的場次以個人發表(一個場次)、Keynote 演講以及 其他個人研究相關者為主。

第一場 Keynote 是由本屆 K. S. Fu Prize 得主 Prof. Horst Bunke (University of Bern, Switzerland)發表的演講，講題為 "Towards the Unification of Structural and Statistical Pattern Recognition" ，試圖將用來處理符號的結構性方法與處理量化資料的統計方法作一統整，相關介紹可至以下連結觀看:

http://www.iapr.org/members/newsletter/Newsletter10-04/index_files/Page679.htm

另一場比較讓我印象深刻的 keynote 是由哥倫比亞大學的 Shree Nayar 教授所演講的 Computational Cameras: Redefining the Image，當中他十分詳盡地介紹了計算攝影學的發展與應用，並舉了很多實例，令我覺得獲益良多。

此外，由 Yahoo!的研究員 Prabhakar Raghavan 所講述的 The Quantitative Analysis of User Behavior Online Data, Models and Algorithms，與本人另一個電子書使用行為紀錄與分析之研究案有相當密切關聯，也是我聽講的重點。其餘參加的場次由於論文數目繁多，因此不再一一詳述。

綜言之，ICPR 是圖型識別領域中相當重要的會議，因此之後也會繼續投稿，參加每兩年一次的盛會。

Region Description Using Extended Local Ternary Patterns

Wen-Hung Liao

Dept. of Computer Science, National Chengchi University, Taipei, Taiwan
whliao@cs.nccu.edu.tw

Abstract

The local binary pattern (LBP) operator is a computationally efficient local texture descriptor and has found many useful applications. However, its sensitivity to noise and the high dimensionality of histogram associated with a mediocre size neighborhood have raised some concerns. In this paper, we attempt to improve the original LBP by proposing a novel extension named extended local ternary pattern (ELTP). We will investigate the characteristics of ELTP in terms of noise sensitivity, discriminability and computational efficiency. Preliminary experimental results have shown better efficacy of ELTP over the original LBP.

1. Introduction

Local binary pattern is a computationally efficient local texture descriptor that has been applied successfully to tasks such as texture classification, face recognition, and background modeling [1]. However, there exist several limitations of LBP that hinder its capability in certain situations. For example, LBP is quite sensitive to random noise in near-uniform image regions. Moreover, LBP with mediocre size of sample points will produce a feature representation with very high dimension. The former issue is, to some extent, resolved by the introduction of local ternary patterns [2]. Yet LTP has a much larger histogram size than the original LBP. The latter problem is generally settled by merging or grouping patterns to reduce the size of the histogram. Yet the dimensionality reduction process usually casts a negative effect on LBP's capability to accurately describe a region. So far, no proposed solution can address both issues simultaneously and effectively.

The objective of this research is to improve the original LBP using a novel extension named extended local ternary pattern. The proposed ELTP accomplishes better tolerance to noise through the incorporation of ternary representation, while at the same time controls the histogram size by merging patterns using a distance measure defined over the ternary digit pattern space.

The rest of this paper is organized as follows. In Section 2 we briefly reviewed previous work, focusing on those directly related to local ternary patterns. Section 3 describes the formulation of extended local ternary patterns, along with some possible variations. Basic properties of the newly proposed ELTPs are also discussed. Section 4 presents some preliminary experimental results and comparative analysis. Section 5 concludes this paper with a conclusion and outlook on future work.

2. Related work

There are various extensions and modifications of the original LBP following its first introduction by Ojala *et al*. [3]. A good source of references can be found in [4]. Since our investigation focuses on issues regarding noise sensitivity and histogram bin size, we will restrict our discussion of related work to these subjects.

According to the original definition of LBP, pattern with a mediocre size of sampling points (P in LBP(P,R)) will generate a histogram of rather high dimensionality. For example, LBP(16,2) will generate a histogram of size 2^{16}=65536, which is not suitable for region description. (A 32x32 image patch will have at most 1024 distinct patterns, resulting in a very sparse representation.) To address this issue, Mäenpää *et al.* [5] proposed two approaches to select a subset of LBP for texture classification. The first method starts with a single pattern and iteratively expands the

pattern collection using a training set. However, the patterns thus chosen tend to depend on the training data employed. The second approach reduces the histogram size by dividing the patterns into uniform and non-uniform ones. The size can be further reduced to *P+2* (*P* is the number of sample points) by incorporating rotation-invariance. Such modifications may prove effective for texture disambiguation, but may face difficulties for tasks such as face or object recognition [6].

The second critical issue regarding LBP is its sensitivity to noise. Let us examine the 3x3 image in Fig. 1. The corresponding LBP is 11001011, or 203. If we change the intensity value of the center-left pixel from 54 to 53, we will obtain a different LBP: 01001011, or 75 (shown in Fig. 2). Notice that these two bit patterns are still quite similar, with their Hamming distance equal to 1. But the values become distinct when converting into decimal representation.
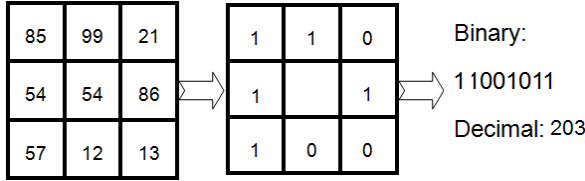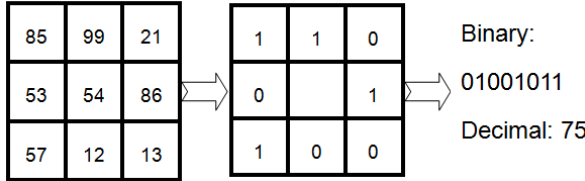

Fig. 1 Calculation of the LBP.

Fig.2 LBP obtained by modifying the value of the center-left pixel from 53 to 53.

.There exist several methods to compute the distance between two histograms. But either histogram intersection (Eq.1) or χ^2 distance (Eq. 2) considers individual bins separately.

$$HI(\boldsymbol{S},\boldsymbol{T}) = \sum_i \min \left\{ \frac{S(i)}{\sum S(j)}, \frac{T(i)}{\sum T(j)} \right\} \qquad (1)$$

$$\chi^2(\boldsymbol{S},\boldsymbol{T}) = \sum_i \frac{(S(i)-T(i))^2}{S(i)+T(i)} \qquad (2)$$

As a result, the slight perturbation caused by replacing a single pixel value yields a rather significant change in pattern distribution and distance measure.

Local ternary pattern seems to be a natural extension of the original LBP to deal with this problem. In [2], Tan *et al.* proposed to use a base-3 pattern to represent the region. The LTP can be calculated according to Eq. (3):

$$LTP(i) = \begin{cases} 1 & \text{if } P(i) - P(0) > \theta \\ 0 & \text{if } |P(i) - P(0)| \le \theta \\ -1 & \text{if } P(i) - P(0) < -\theta \end{cases} \qquad (3)$$

where *P(0)* is the intensity of the center pixel, and θ is a pre-defined threshold. Using this new representation, the two image patches in Figs. 1 and 2 can both be converted to the same ternary pattern shown in Fig.3.
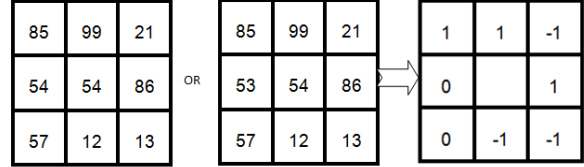

Fig. 3 LTP obtained by setting θ=5.

While this approach seems to have addressed the noise sensitivity issue, it actually creates another problem regarding the histogram dimensionality. For example, *LTP(8,1)* will generate a histogram of size 3^8=6551. *LTP(16,2)* will create a histogram of size 3^{16}=43,046,72, which is deemed inappropriate for practical implementation.

Dimensionality reduction is achieved in [2] by decomposing the LTP into two LBPs (upper and lower), as demonstrated in Fig. 4. The distance between two LTP is calculated by combing the results from upper and lower LBP, respectively.
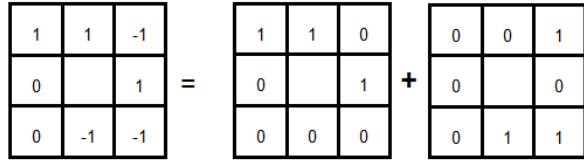

Fig. 4 Decomposing LTP into two LBPs.

The above coding scheme has brought relief to the histogram dimensionality problem. However, it also adversely affects the pattern's tolerance for noise, which is considered to be the key strength of LTP. According to our experiments, LTP-UL-LBP performs even worse than the original LBP in the presence of noise. An effective coding scheme needs to be developed to maintain compactness of the feature vector, while at the same time retain its ability to faithfully describe the object of interest in a noisy environment.

3. Extended Local Ternary Patterns

There are two conflicting factors that affect the performance of LBP/LTP. On one hand, when we

employ a larger number of sample points or use a base-3 representation, we will achieve better feature resolution. One the other hand, too fine the resolution implies sensitivity to minor changes in the pattern, as well as difficulties in actual implementation. (Consider the case of LTP(16,2)). The proposed ELTP attempts to strike a balance by using a clustering method to group the patterns in a meaningful way, as explained in the following.

3.1. Basic formulation

The process for converting a region into its ELTP representation (Eq. (4)) is very similar to that described in [2], i.e.,

$$LTP(i) = \begin{cases} 1 & \text{if } P(i) - P(0) > \theta \\ 0 & \text{if } |P(i) - P(0)| \leq \theta \\ -1 & \text{if } P(i) - P(0) < -\theta \end{cases} \quad (4)$$

Instead of employing a fixed threshold θ, however, we propose to assign its value based on the local statistics of the pattern. Specifically, we will use Eq. (5) to compute θ:

$$\theta = \alpha \times \sigma \quad (0 < \alpha \leq 1) \quad (5)$$

where σ is the standard deviation of the local patch, and α is a scaling factor. Such a formulation helps to retain one favorable property of LBP: invariance with respect to illumination transformation, as illustrated in the following example

Fig. 5 depicts the LTP of two image patches using our proposed criteria (α=0.3). The intensity value of the right patch is obtained via a simple linear transform:

$$R(i) = L(i) \times 3 + 10 \quad (6)$$

10	40	45
23	25	23
40	15	35

or

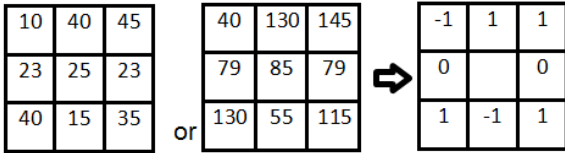| 40 | 130 | 145 |
| 79 | 85 | 79 |
| 130 | 55 | 115 |

⇨

-1	1	1
0		0
1	-1	1

Fig. 5 Invariance of ELTP under gray-level transformation.

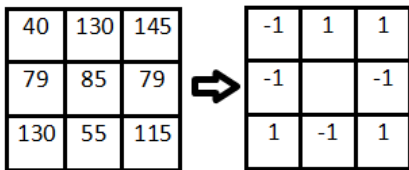If we use fixed threshold, say θ=5, the right region will have a different LTP, as shown in Fig. 6.

40	130	145
79	85	79
130	55	115

⇨

-1	1	1
-1		-1
1	-1	1

Fig. 6 LTP using a fixed threshold (θ=5)

3.2. Dimensionality reduction

As discussed previously, using a base-3 system for representing feature patterns will increase the feature dimension in a drastic manner. It is therefore necessary cut down the size of the histogram by grouping patterns. But how does one achieve this goal in a sensible manner? Here we propose to form the groups based on pattern similarity. Suppose $x = (xn-1,...,x0)$ and $y=(yn-1,...,y0)$ are two ELTP strings, the distance between x and y can be calculated using their Hamming distance:

$$D(x,y) = \sum_{i=0}^{n-1} |x_i - y_i| \quad (6)$$

For a ternary string of length n,

$$\max \quad D(x,y) = 2n \quad (7)$$

The similarity (or affinity) between two ELTP strings can therefore be defined as:

$$A(x,y) = 1 - \frac{D(x,y)}{2n} \quad (8)$$

When there is a need to group patterns, those with larger affinity should be merged together. Specifically, if P is the number of sample points, there will be at most 3^P distinct ELTP strings. To reduce the size of the feature dimension from 3^P to K, we will first compute the similarity between any two ELTP strings to form a $3^P \times 3^P$ symmetric affinity matrix. This will transform the original dimensionality reduction problem into a graph partitioning problem, which can be solved using spectral clustering algorithms [7]. It should be noted that the same process can be applied to reduce the feature dimension of LBP. The only difference lies in the way one computes the similarity measure (Eq. 9):

$$A_2(x,y) = 1 - \frac{D(x,y)}{n} \quad (9)$$

Fig. 7 summarizes the procedure for reducing feature dimension for the proposed ELTP representation.

S1. Choose P(sample points) and K(histogram size)
S2. Form a $3^P \times 3^P$ affinity matrix using Eq. (8)
S3. Perform a K-way partition of the 3^P patterns using spectral clustering.
S4. Merge those patterns belonging to the same partition into a single bin in the histogram.
S5. Use the K-dimensional histogram for feature representation.

Fig. 7 Dimensionality reduction process for ELTP.

Following the K-way partitioning, it is possible to compute the mean distance of patterns belonging to the same cluster. The mean distance is regarded as an indicator of cluster homogeneity, and can be used to assign the weight w_i in weighted χ^2 distance (Eq. 10). Clusters of greater homogeneity should get more weight.

$$\chi_w^2(\boldsymbol{S}, \boldsymbol{T}) = \sum_i w_i \frac{(S(i)-T(i))^2}{S(i)+T(i)} \qquad (10)$$

The uniform pattern argument also applies in the ternary case, with a slight modification in the definition. If we consider only the number of transitions in the ternary digit pattern, then both 12222111($1\rightarrow2$, $2\rightarrow1$) and 02220000 ($0\rightarrow2$, $2\rightarrow0$) will be uniform patterns. However, if we take into account the amount of transitions, then the former pattern (transition amount=2) will be more uniform than the latter (transition amount=4).

Finally, ration-invariance version of the ELTP can be obtained by defining a new distance measure (Eq. 11):

$$D'(x, y) = \min_{j=0\ldots n-1} D(x, ROT_j(y)) \qquad (11)$$

where ROT_j denotes the circular shift operation by j digits.

4. Experimental Results

We present preliminary experimental results comparing the performance of the original LPB, LTP-UL-LBP and the newly proposed ELTP in term of noise sensitivity. For the purpose of comparison, we set $P=8$ and $K=256$. The scaling factor α for defining ELTP is set to 0.3 in the experiment. We use lena image to perform the test. The image is corrupted with Gaussian noise of different scales. Image patches are randomly selected from the noisy image and the corresponding LBP, LTP and ELTP are calculated. To evaluate noise immunity, we compute the histogram intersection between the original patterns and their noisy counterparts. The results are depicted in Fig. 8.

Generally speaking, ELTP is least sensitive to perturbations, especially at high noise levels, using the same size of feature vector ($K=256$). The original LTP was designed to have better noise resistivity, yet the coding scheme (by decomposing into upper and lower LBP) counters all the benefits. As for the computational complexity, the K-way partition needs to be performed only once. After that, the grouping of patterns can be done using a fairly simple table-lookup method.
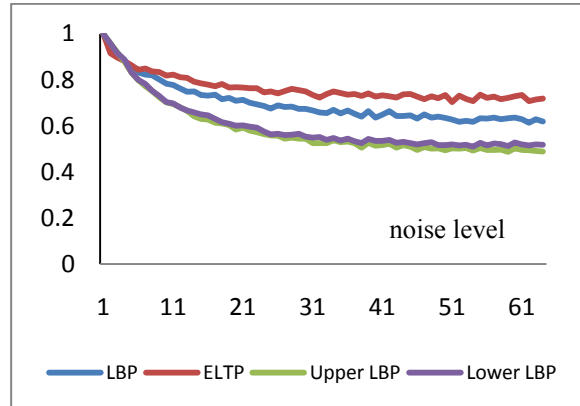

Fig. 8. Performance comparison.

5. Conclusions

A novel scheme of defining local ternary patterns and a systematic approach for grouping these patterns have been devised in this paper. Preliminary experimental analysis showed encouraging results using the proposed ELTP for region description.

Future work includes an in-depth investigation of different spectral clustering algorithms and how they affect the partitioning results. More importantly, we will examine the efficacy of the proposed ELTP to machine vision applications such as texture classification, face or facial expression recognition and background modeling.

References

[1] T. Mäenpää and M. Pietikäinen, Texture Analysis with Local Binary Patterns. In: Chen CH & Wang PSP (eds) *Handbook of Pattern Recognition and Computer Vision*, 3rd ed, World Scientific, pp. 197-216, 2005.

[2] X. Tan and B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions". In *Analysis and Modeling of Faces and Gestures*, volume 4778 of LNCS, pp.168–182. Springer, 2007.

[3] T. Ojala, M. Pietikäinen, and D. Harwood, A Comparative Study of Texture Measures with Classification Based on Feature Distributions, *Pattern Recognition,* vol. 29, pp. 51-59,1996.

[4] http://www.ee.oulu.fi/mvg/page/lbp_bibliography

[5] T. Mäenpää, T. Ojala, M. Pietikäinen and M. Soriano, Robust Texture Classification by Subsets of Local Binary Patterns. *Proc. 15th International Conference on Pattern Recognition,* Vol. 3:pp. 947-950, 2000.

[6] T. Ojala, M. Pietikainen, and T. Maenpaa, Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002

[7] U. von Luxburg, A Tutorial on Spectral Clustering. *Statistics and Computing*, Vol. 17(4) pp. 395-416, 2007.

國科會補助專題研究計畫項下出席國際學術會議心得報告

計畫編號	NSC 99－2221－E－004－012－
計畫名稱	基於延展式區域三元化樣式的物件描述方法

出國人員 姓名	廖文宏	服務機構 及職稱	國立政治大學資訊科學系
會議時間	100 年 7 月 6 日至 100 年 7 月 8 日	會議地點	美國喬治亞州 Athens

會議名稱	(中文)第十一屆 IEEE 國際學習科技研討會 (英文) 11th IEEE International Conference on Advanced Learning Technologies
發表論文 題目	(中文)電子書閱讀器使用者行為資料分析：以高中生為例 (英文) Analysis and Interpretation of e-Reader User Logs: A Case Study of High School Students' User Behaviors

The 11th IEEE International Conference on Advanced Learning Technologies (ICALT 201)論文接受函

Dear Wen-Hung Liao

Thank you for your paper submission to the 2011 IEEE International Conference on Advanced Learning Technologies (ICALT 2011).

The review process is now completed. This year a large number of submissions were received and the overall quality of the submitted papers was very high.

In some instances we had to make a decision based not only on the paper quality but also on how well the work fit with the main themes of the conference.

We are pleased to inform you that your paper:

Paper ID: **145**

Title: **Analysis and Interpretation of e-Reader User Logs: A Case Study of High School Students' User Behaviors**

has been ACCEPTED as: **SHORT PAPER**, with a page limit for your final camera ready manuscript of 3 pages.

Please note that, according to the recently adopted **IEEE Conferences No Show Policy, we should notify you that** *"IEEE reserves the right to exclude a paper from distribution after the conference (e.g., removal from IEEE Xplore) if the paper is not presented at the conference"*. Therefore, please take into serious consideration this notification when planning your participation to the ICALT 2011 Conference.

Guidelines and important dates for both the submission of the Camera-Ready Manuscript and the IEEE Copyright Form, and the Authors' Registration will be available on the ICALT 2011 Conference Web Site. Please also note that you will receive an e-mail from IEEE Conference Publishing Services (CPS), after your registration and payment to the ICALT 2011 Conference, containing an author-kit for preparing and submitting your camera-ready paper

Thank you for your contribution to ICALT 2011.

Looking forward to meeting you in Athens, Georgia, USA this summer!

Best regards,

Ignacio Aedo, Universidad Carlos III de Madrid, Spain
Nian-Shing Chen, National Sun Yat-sen University, Taiwan
Demetrios Sampson, University of Piraeus & CERTH, Greece

ICALT 2011 Program Co-Chairs

ICALT 2011 心得報告

IEEE International Conference on Advanced Learning Technologies 是與數位學習領域相關的研討會，ICALT 2011 於 7/6-7/8 在美國喬治亞州喬治亞大學所在地 Athens 舉行，因為執行校園用電子書專案之緣故，原來對此研討會較為陌生的我，藉由投稿 **Analysis and Interpretation of e-Reader User Logs: A Case Study of High School Students' User Behaviors** 的機會，第一次參與數位學習相關的國際會議，由於經驗較缺乏，因此抱持學習的成分居多，也對不同領域學者的觀點與研究取徑有一番觀察與比較。

此次研討會的主題為:**Educational Technology Forecast: Cloudy with a Slight Chance of Gain**，有了科技輔助，是否未來學習就一片萬里晴空，或者仍然多雲陣雨?顯然在數位化工具越來越普遍的今日，資訊科技究竟能在教育學習扮演何種積極之角色，有再度審視與省思的空間，研究學者不再以開發新平台或推動新工具為唯一目標，而是更客觀、忠實地評估這些工具的正面與負面衝擊，這種勇於反思的精神，正是所有研究人員應該抱持的態度。

ICALT 2011 在喬治亞大學(University of Georgia)的會議中心舉行，該大學創校於 1785 年，至今已有兩百多年的歷史，因此校園內建築頗為多樣化，有新穎的體育場，也有具歷史的古老建物，漫步於校園四周，令人發思古之幽情。

在會議的安排部分，除了每天早上的 Keynote 或 invited speech 外，共分四個平行場次，每天安排到晚上六點結束，poster 部分從下午三點才開始，因此整個會議 oral presentation 的比重較高，至於論文報告的安排，則依照下列主題分組進行。從下面的主題列表，也可略窺目前數位學習領域研究的若干重點。

- Adaptive and Personalized Technology-enhanced Learning I,II,III,IV
- Learning Systems Platforms and Architectures I,II,III,IV
- Semantic Web and Ontologies for Learning Systems I,II
- Computer Supported Collaborative Learning I,II
- Digital Game and Intelligent Toy Enhanced Learning I,II,III
- Wireless, Mobile and Ubiquitous Technologies for Learning I,II
- Virtual Worlds for Academic, Organizational, and Lifelong Learning I,II
- Web 2.0 and Social Computing for Learning and Knowledge Sharing I, II
- Technology-enhanced Science Education I,II
- Rethinking Pedagogy in Technology-enhanced Learning I,II
- e-Assessment and new Assessment Theories and Methodologies
- Knowledge and Competencies Management
- Emerging technologies for enhancing learning processes and outcomes, and diagnostic of mental representations in learners
- Open Content and Learning Technology Standards
- Self-Regulated Learning in Responsive Open Learning Environments

- Affective and Pervasive Computing for Learning

我們的論文被排在 7/7（周四）Parallel session 4a (11:00-13:30) Wireless, Mobile and Ubiquitous Technologies for Learning II 進行口頭報告，由於我們進行研究所使用的電子書系統是自行開發，有關系統整合、介面設計、教材轉置均是由本團隊所執行，因此除了分享高中生使用行為之觀察與 log 分析外，也帶了幾台機器供現場聽眾試用，以提升互動效果，當場有三位聽眾提出問題，多能切中核心，並未有太令人意外的疑問發生。

除了早上的公開演講，在口頭報告部分，我挑選的場次仍與技術較有密切相關者為主，如 Digital Game and Intelligent Toy Enhanced Learning，Wireless, Mobile and Ubiquitous Technologies for Learning，Adaptive and Personalized Technology-enhanced Learning 等，有些論文有完整的論述，從學習理論出發，搭配設計方法與實測結果及分析，有些則屬較前期的開發評估，概念新穎但效果仍待驗證，從這些不同學科或方法探討中，讓我對數位學習有了更廣闊的見識。

此次參與會議的台灣學者也不少，有台科大、台師大、中央、成大、中山等研究團隊，在國科會的領域分類屬於科教處，過去本人接觸不多，藉這次會議得以結識更多學者，也是額外的收穫。

Analysis and Interpretation of e-Reader User Logs

A Case Study of High School Students' User Behaviors

Wen-Hung Liao, Chien-Pao Chueh

Dept. of Computer Science
National Chengchi University
Taipei, TAIWAN
e-mail: whliao@cs.nccu.edu.tw

Abstract—**This study explores the daily life user experiences of an experimental e-book reading device among high-school students, aiming to understand how well the digital natives accept the use of e-book reading devices and the potential utilities of such devices for them, either for leisure purposes or as an assistive educational tool. Toward this goal, we have custom-designed the e-reader user interface as well as the e-book content to suit the needs of this particular user group. The unique opportunity of having access to the hardware device, software design and potential users creates an ideal experimental platform for us to unbiasedly investigate the role of this new technology through a long-term user behavior collection and analysis process. We anticipate that the new reading behaviors of the digital natives will provide clues for further improvements in the design and development of digital reader devices.**

Keywords- e-ink display; e-book reader; user log analysis; user experience

I. INTRODUCTION

Digital natives, the generation growing up with abundance digital technologies, have developed new ways of obtaining and consuming information. With continuing advances in information technology, especially in the area of display hardware, market for the once slow growing e-book reading devices has revived. The introduction of Kindle in 2007 by Amazon has kick-started a new wave of devices based on electronic-ink display. All the major book retailers in the U.S. have followed suit, aiming to replicate the success of Amazon.

The format of digital electronic books has been in existence for decades. The recent surge of e-book readers is mostly attributed to the reduced manufacturing cost of e-ink display. Unlike the conventional LCD, the e-ink screens reflect lights like ordinary paper. It causes less eye strain, and consumes very little power. Anecdotal evidence suggests that e-paper is more comfortable to read than conventional displays. However, current e-ink devices have only 16 levels of gray. The refresh rate is also too slow to show animation or video. These constraints have led the device makers to emphasize its portability, readability and 'greenness'.

Although many e-book readers have been introduced in the last two years, they are mainly targeted on general users for their casual reading activities. Several field studies have been conducted to gather the initial responses of the users to this new technology [1,2]. A few pilot programs sponsored by Amazon have been carried out in several U.S. universities to study Kindle DX's functionality and potential to replace paper textbooks in college. The initial responses are disapproving, due to the device's slow performance and lack of note taking capability [3].

In this research, we are aiming at a different user group: high school students. We also have a different perspective. Instead of forcing the participant to use the device according to our direction, the objective is to exploit the role of this device in the student's daily lifer user experiences through a long-term user behavior collection and analysis process. We are able to obtain the hardware from a supportive manufacturer. The device is based on a proprietary embedded OS. With access to the software development kit (SDK), we have the capability to customize the e-reader user interface, as well as the system functionalities of the device. To understand how this device is used, we asked the participants to fill out surveys. Additionally, a background process is embedded in the device to gather user log for subsequent analysis.

The rest of this paper is organized as follows. In Section 2 we list the specifications of the device and discuss some design constraints. We then present the system framework and user interface design. Section 3 is concerned with the experiment design, including subject selection, deployment process and test schedule. Section 4 presents the user log analysis. Section 5 concludes this paper with a brief summary and outlook on future work.

II. DEVICE AND USER INTERFACE DESIGN

The e-book reading device used in this experiment has dual screens, as depicted in Fig. 1. The top screen is a six-inch e-ink display panel with a 600x800 resolution and 16 levels of grayscale. It is used for displaying information. The bottom is a LCD-based touch screen whose resolution is 360x100. This area is dedicated for user operation and interaction. The main screen shows the six functions of the e-book reading device. My textbooks item (A) contains course materials and assigned readings. Recent readings item (B) provides a short cut to recently opened documents. My documents item (C) allows the user to access self upload pdf or word doc files. My music item (D) contains audio collections. My pictures item (E) is a place holder for png, jpg or gif files. System settings item (F) permits the

user to adjust the device's settings, including data/time, refresh mode, etc.

Due to limited display area and number of gray levels, not all textbook materials are suitable to be presented using this device. Four subjects, namely, Chinese literature, English, History and Geography have been identified after consultation with high school teachers.
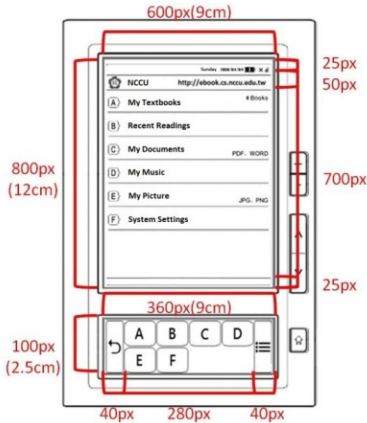

Figure 1. Specifications and user interface of the e-reader.

III.   EXPERIMENT DESIGN

Forty 10th-grade students from the affiliated high school of National Chengchi University (AHS of NCCU) participated in the experiment after obtaining written consent from their parents. We held a discussion session to solicit input from the potential users during the design stage. We have carefully described the characteristics of the e-ink display for the users to better understand the limitations of this device. The users are encouraged, yet not forced, to use the device as much as they can.

To understand how this device is used, we ran a log-collection process in the background to record user activity. Specifically, all the key-press events, including hardware and virtual keys, the time stamp, and the corresponding page will be written to a log file. The format is shown in Table 1.

TABLE 1. FORMAT OF THE USER LOG

	Key Pressed	Time	Page
1	VK_Menu	1272888558	Page_ResourcesCenter
2	VK_A	1272888562	Page_MiniReader

The first phase of our experiment starts at 5/5/2010. We gathered user logs at 5/31/2010, a little more than three weeks after the initial distribution. Out of the forty machines sent out to the students, thirty-eight log files have been successfully collected. Two devices were sent back for repair due to fragility of the display panel.

The second phase of our experiment falls between 6/1/2010 and 7/2/2010, right before the summer vacation starts. For this period, we are able to recover only 34 system logs since four more devices were out of order and sent for repair. We made our final collection of machine logs on 10/29/2010, almost 6 months after the initial distribution. We are able to record 33 user logs this time.

IV.   USER LOG ANALYSIS

This section discusses the results of user log analysis. We will first present the overall usage statistics and identify active users. We will then inspect the operation log of active users to study how the e-book readers have been used in a real-life situation.

A. Overall Statistics

Fig. 2 summarizes the usage statistics of the thirty-three students who participated in all three phases of the experiment. About 50% (17 out of 33) of the users spent less than 30 hours on the machine. Our speculation is that these are individuals who may be curious about the device, but the initial excitement wears down quickly after hands-on experience. These students are classified as 'inactive' users. Follow-up questionnaire has been distributed to inquire the reasons for low utility. Ten participants whose usage ranges from 30-60 hours are classified as 'casual' users. The remaining six participants whose usage exceeds 60 hours are considered 'active' users.

The next useful information is to examine the average time between consecutive key-presses (or clicks) to deduce the possible user behavior. If the interval between two operations is short, the user may just be navigating through different menu structures. On the other hand, if the interval is long, we have reason to believe that the user is reading the material shown on the e-ink display. The average time between two operations (T_{cc}) can be computed by:

$$T_{CC} = \frac{Total\ Usage\ in\ Seconds}{Total\ \#\ of\ clicks} \tag{1}$$

Fig. 3 depicts the calculated T_{cc} for all 33 users. The average is 26.1 seconds with a standard deviation of 9 seconds.
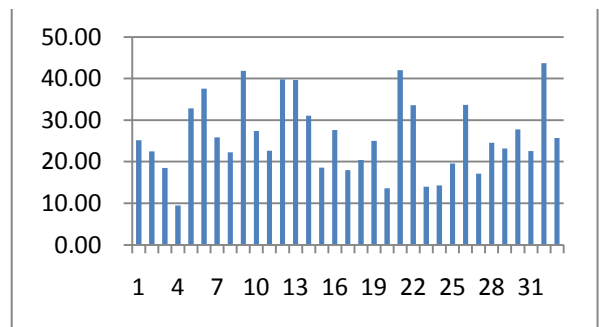

Figure 3. Average time (in seconds) between two operations

We speculate that active users spend most of their time reading the e-books instead of jumping around different hierarchies of user menus. Consequently, their T_{cc} should be longer than that of other types of users. This is confirmed by the statistics shown in Fig. 4, in which all 6 active users have T_{cc} greater than 30 seconds. For inactive or casual

users, the distinction is not as clear, as can be seen from the mixed results in Fig. 4.
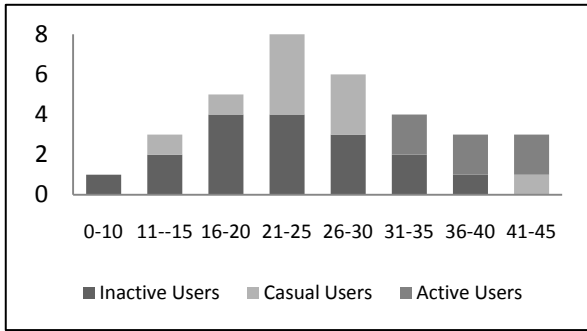

Figure 4. Histogram of T_{cc}

B. Statistics of Active Users

Certain statistics only make sense for those logs obtained from active users. As a result, the subsequent analysis will focus on the detailed usage patterns of the active users.

Fig. 5 shows the percentage of time active users spent on different type of materials. Either textbooks or assigned readings (i.e., classic Chinese or English novels) are not very well-received, totaling only 25% of the overall reading time. Instead, most active users prefer to upload their own documents. We believe this is strongly correlated to the policy we have in this phase of the experiment, i.e., we treat the e-book reader as a complementary tool and let the user decide when, where, and how to utilize the machine.

A user session is defined as the interval between system boot-up and standby or shut down. We can tell if the e-book reader is used in fragmented time intervals by examining the statistics of the user session. If we set the threshold to be 300 seconds (5 minutes), the average is 34.7% of long sessions (over 5 minutes) vs. 65.3% of short sessions (less than 5 minutes) for all active users.
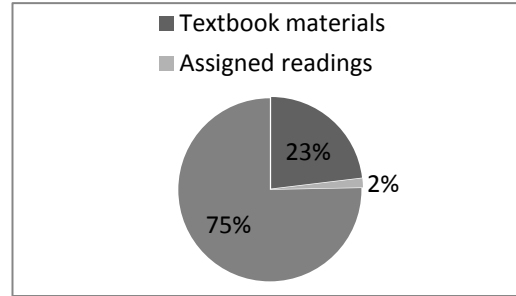

Figure 5. Usage pattern: type of materials consumed

V. CONCLUSIONS

In this paper, we have described an experiment aimed at facilitating high school learning through e-book reading devices. Results from user log analysis indicate that the reader can play an assistive role in promoting reading activities. We are currently entering the second stage of design and experiment using a new hardware platform running Android OS. We are attempting to extend the e-pub format to enable the incorporation of online tests and interactive content. It is expected that with these new functionalities and integrated online services, the e-book readers will gain wider acceptance in modern classrooms.

REFERENCES

[1] Dennis T. Clark, Susan P. Goodwin, Todd Samuelson, and Catherine Coker, 2008. A qualitative assessment of the Kindle e-book reader: results from initial focus groups, Performance Measurement and Metrics, Vol. 9 Issue: 2, pp.118 – 129, 2008.

[2] Guy Janssens and Harry Martin, 2009. The feasibility of E-Ink readers in distance learning: A field study, International Journal of Interactive Mobile Technologies, Vol. 3, No. 3, pp. 34-37, 2009.

[3] The Kindle DX pilot project at Darden, http://www.darden.virginia.edu/html/standard.aspx?menu_id=72&styleid=2&id=19304
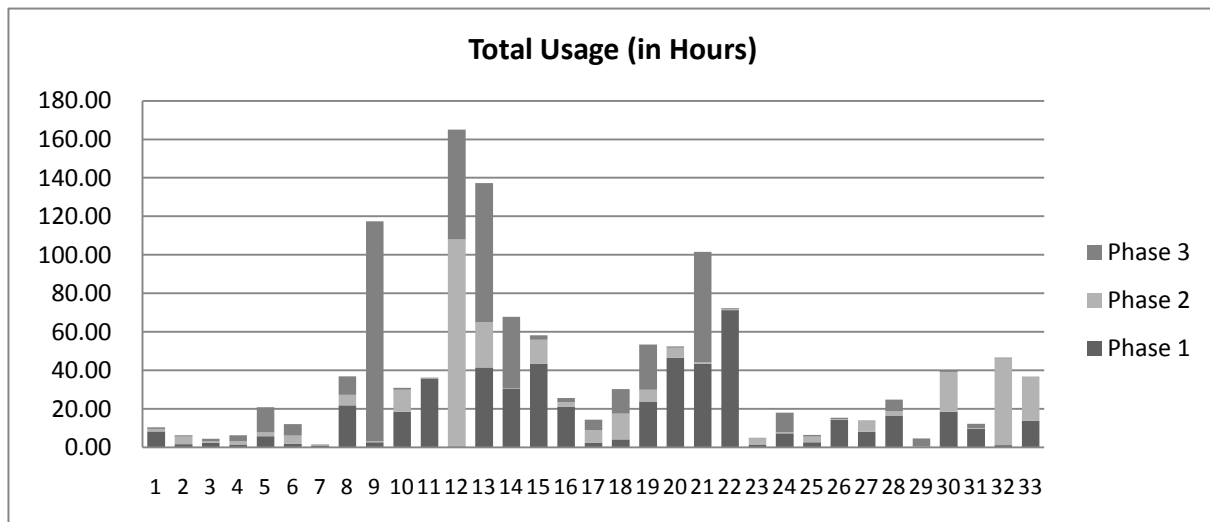
Figure 2. Total usage (in hours) in three experiment phases

國科會補助計畫衍生研發成果推廣資料表

國科會補助計畫	計畫名稱: 基於延展式區域三元化樣式的物件描述方法
	計畫主持人: 廖文宏
	計畫編號: 99-2221-E-004-012-　　　　學門領域: 圖形辨識

<div align="center">無研發成果推廣資料</div>

99 年度專題研究計畫研究成果彙整表

計畫主持人：廖文宏　　計畫編號：99-2221-E-004-012-

計畫名稱：基於延展式區域三元化樣式的物件描述方法

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	2	2	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	1	1	100%	人次	
		博士生	1	1	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		

	成果項目	量化	名稱或內容性質簡述
其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等,請以文字敘述填列。)	無		

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

---

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估
   ■達成目標
   □未達成目標（請說明，以 100 字為限）
   　　□實驗失敗
   　　□因故實驗中斷
   　　□其他原因
   說明：

---

2. 研究成果在學術期刊發表或申請專利等情形：
   論文：■已發表 □未發表之文稿 □撰寫中 □無
   專利：□已獲得 □申請中 ■無
   技轉：□已技轉 □洽談中 ■無
   其他：（以 100 字為限）
   發表論文
   1. W. Liao, 'Region Description Using Extended Local Ternary Patterns', Proceedings of the 20th International Conference on Pattern Recognition, pp. 1003-1006, 2010.
   2. W. Liao and T. J. Young, 'Texture Classification Using Uniform Extended Local Ternary Patterns', Proceedings of The IEEE International Symposium on Multimedia, pp. 191-195, 2010.

---

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限）

   本研究所提出之 extended local ternary pattern，乃基於著名的區域描述法 local binary pattern，期望在不增加太多計算量的前提之下，改善原有方法之抗噪力，使其適用於各式之光源條件與運作環境。

   三元化的想法，並非本研究肇始，然而先前提出的定義方式，只有點出問題，而沒有真正的解決問題，關鍵的降維概念並未被導入，以致於只能紙上談兵，甚或使用了錯誤的降維方式，反而使得整體的效能其準確度都降低。

   本研究使用 spectral clustering 的分群方式，作為降維之手段，不但具有通用性，也有應用上之彈性，此外，本研究亦探討這類型描述子的重要的性質:uniform pattern，在新的定義基礎上，重新檢視 uniform pattern 的組成與分布，透過這個面向的研究，豐富了未來發展的可能性。

   以學術貢獻而論，目前本研究之成果已發表於兩個國際知名會議: Internal Conference on Pattern Recognition（ICPR 2010），International Symposium on Multimedia（ISM

2010），提出之立論獲 reviewer 的高度贊同，後續的一些成果正在整理當中，將有更多的發表出現。

以技術創新的層次而言，ELTP 要解決的是電腦視覺中物件描述與辨識的問題，在實務上有許多重要的應用。從本研究的實驗數據顯示，不管是材質分類、人臉辨識，甚或背景建模等，ELTP 均有著優於 LBP 的表現，因此將可有效作為物件辨識核心的主要特徵描述方法。就未來的發展而言，因 ELTP 乃基於 LBP，而 LBP 本身也有許多的變形，如 Rotation-invariant LBP、XY-LBP、volume LBP（VLBP）、CS-LBP 等，如何將本研究所提出三元編碼的概念套用至上述的描述子中，將是值得後續研究探討的問題。