適用於預測差值直方圖位移藏密法之藏密分析技術

劉興漢[1]　劉江龍[2]

[1]國防大學管理學院資訊管理學系
[2]國防大學理工學院電機電子系

論文編號：3602-6
收稿 2015 年 07 月 29 日 → 第一次修訂 2015 年 8 月 27 日 → 同意刊登 2015 年 10 月 08 日

摘要

自從 911 事件發生後，資訊隱藏及偵測技術成為各國國防安全上的熱門研究課題。預測差值直方圖位移藏密法(PEHS)可有效的在影像中藏入大量秘密訊息，但其在預測差值的直方圖上卻有不正常的分佈。本論文提出可有效偵測 PEHS 藏密影像之偵密技術，主要是透過預測差值直方圖的比率來進行特徵值的擷取，並利用倒傳遞類神經網路進行分類，以分辨是否為使用 PEHS 之藏密影像。實驗結果證明，本文所提出之偵測技術對 PEHS 偵測正確率達 99%以上，與 2007 年 Zhang 學者與 2010 年 Pevný 學者提出之偵測方法所提供之 69.7%與 70.9%偵測率比較，可有效提升對 PEHS 藏密影像的偵測正確率，適合國防安全上之應用。

關鍵詞：資訊隱藏，資訊隱藏分析技術，預測差值直方圖

Specific Steganalysis for Detection Prediction-Error Histogram Shift Steganography

Hsing-Han Liu [1]　Chiang-Lung Liu [2]

[1] Department of Information Management, Management College, NDU, Taiwan, R.O.C.
[2] Department of Electrical and Electronic Engineering, CCIT, NDU, Taiwan, R.O.C.

Abstract

Information hiding and steanalysis have become hot research topics in the field of national defense security after the 911 tragedy. The Prediction-Error Histogram Shift method (PEHS) can effectively hide large amount secret message in an image. However, the histogram of the Prediction-Error differences appears abnormal distribution. In this paper, we propose a steganalysis method which can effectively detect the stego images created by PEHS. The proposed method uses the ratio of Prediction-Error to extract the image feature. The image feature is then classified by Back-Propagating Neural Network. Experimental results show that the accurate detection rate of the proposed steganalysis method to PEHS is above 99% and outperform the performance of 69.7% and 70.9% provided by the method proposed by Zhang et al. in 2007 and Pevný et al. in 2010. Therefore, the proposed steganalysis method can effectively improve the detection rate for detecting the stego images created by PEHS and is practical for applications of national defense security.

Keywords: Steganography, Steganalysis, Prediction-Error Histogram.

壹、前言

由於電腦及相關資訊產業技術的蓬勃發展，加上網際網路運用越來越生活化，使得以往需要依靠人力的資料傳遞方式，轉變成只需透過網路即可進行數位資料的交換及流通。但這如此便利的環境就如同雙面刃，若有不法人士企圖利用此環境竊取機密資訊，對個人或團體所造成的損失將難以估計。為了保護個人或團體的機密資訊，可透過密碼學(Cryptograghy)的資料加密技術(Paar and Pelel, 2010)，以確保其機密性(Confidentiality)。但由於資料加密技術會針對原始資料內容進行處理，使其成為不具任何意義的亂碼，反而成為加密文件的重要象徵，引導有心人士針對此加密文件加以破解或破壞，造成加密後資訊在安全上的另一項隱憂。為了解決此問題，資訊隱藏技術(Information Hiding)因此興起。

資訊隱藏可將機密資訊嵌入於特定媒體內，產生隱含機密的影像或聲音等載體(Carrier)，以避免機密被有心人士所發覺，進而達到防止機密資訊遭到破壞或竊取之目的。資訊隱藏技術依其運用目的之不同，可區分為兩種重要類型。第一種是為了確保影像完整性所發展出的數位浮水印技術(Watermarking)，其是在不影響載體完整性(Integrity)下，進行訊息的嵌入(Embedding)，達到保護智慧財產權的目的；第二種是為進行秘密通訊所發展出來的藏密技術(Steganography) (Petitcolas et al., 1999)，其主要作法是將欲傳遞的秘密訊息嵌入掩護載體(Cover Carrier)中，以躲避第三者的察覺，而合法接收者在取得已藏密的偽裝載體(Stego Carrier)後，則可依取密程序取出秘密訊息。

網際網路的高度發展與資訊科技的進步提供了藏密技術一個良好的發展環境，而數位影像則因具有大量散佈、容易取得、及有龐大的藏密空間等特性，已成為最普遍使用的掩護載體(Rabah, 2004)。一般而言，好的影像藏密技術須能符合不可察覺性(Imperceptibility)與高資料負載(Payload)的要求(Cacciaguerra and Ferretti, 2000)。所謂不可察覺性，是指被嵌入秘密訊息之藏密影像(Stego Image)不可顯露出人眼或統計上可察覺到的人工失真(Artificial Distortion)；而高資料負載則是為了滿足傳輸大量秘密訊息之需求。密碼學與藏密技術在秘密通訊之實務上屬相輔相成，為增加秘密通訊的安全性，欲傳送的秘密訊息常經過密碼學技術加密後，再嵌入載體影像(Cover Image)後傳送，其關係如圖 1所示。

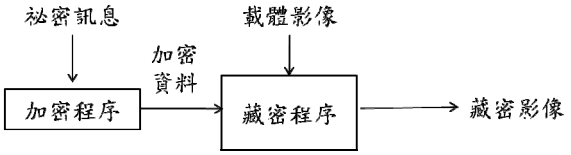

圖1　密碼學與藏密技術之關係圖

最簡單的空間域藏密方式是利用影像像素的最低位元(Least Significant Bit, LSB)來嵌入秘密訊息，又稱 LSB 取代藏密法，簡稱為 LSB 藏密法。Bender(1996)等學者首先提出 LSB 藏密法，其優點為人類視覺無法察覺秘密訊息的存在，而其缺點為容易被 RS 偵測技術(Fridrich et al., 2001)及 x^2 攻擊法(Westfeld et al., 1999)等統計式的分析技術所攻擊。為了改善此缺點，許多學者提出許多以 LSB 為基礎的藏密技術(Mielikeainen, 2006)。

藏密技術依據在取密後是否能完全回復載體影像的特性，可區分為失真與無失真藏密技術。失真藏密技術適用於不需確保載體影像完整性之應用，例如上述 LSB 藏密法屬於此類；而對於醫學及軍事上的特殊應用，需要在秘密訊息取出後仍可確保載體影像的完整性，則適用於無失真的藏密技術。無失真藏密技術大致可分為三大類(Feng, 2006)，第一類是利用資料壓縮的技術對影像特徵值修改以進行藏密；第二類是利用可逆的像素值運算方法進行藏密；第三類是利用影像像素統計直方圖位移(Histogram Shifting, HS)進行藏密。由於直方圖位移法的無失真藏密技術對於影像造成的失真最小，藏密後的影像品質較高，吸引非常多的學者進行研究。

2006 年，Ni 等學者(2006)提出 HS 無

失真藏密技術（簡稱 HS 藏密法），其方法是先統計載體影像像素值的分佈，將其繪製成直方圖，再藉由平移峰值點(Peak Point)與零值點(Zero Point)之間的像素值一個單位（加 1 或減 1）的方法，以嵌入秘密訊息。雖然 HS 藏密法所產生的藏密影像品質良好，但其藏密量不高，因此陸續有學者提出改進 HS 藏密法的藏密技術。例如 Hwang 等學者(2006)同時利用載體影像直方圖中兩個峰值點（峰值點加 1 與減 1 的位置）進行秘密訊息的藏入，藉此提升藏密量；Fallahpour 等學者(2007)提出將影像分割成不重疊的多個區塊，再分別進行 HS 法藏密，以較多組的峰值點來提高藏密量；Tseng 等學者(2008)則進一步將載體影像分割成 4×4 個不重疊影像區塊，而每一分割區塊運用 2 組峰值點進行藏密，以增加藏密量。

結合預測差值(Prediction Error, PE)與 HS 的 PEHS 藏密技術是目前無失真藏密技術最新發展趨勢。PE 為載體影像的像素值與經由預測器(Predictor)產生的預測像素值之差值。由於 PE 具有大量接近 0 值的特性，可在預測差值直方圖(Prediction Error Histogram, PEH)產生較高的峰值點統計量。與基於 HS 的藏密技術相較，基於 PEHS 藏密技術的藏密量明顯較高，例如 Hong 等學者(2009)、Tsai 等學者(2009)、及 Kim 等學者(2009)所提出之藏密技術，均為此例。

雖然藏密技術可確保機密資訊的安全，但恐怖組織或電腦罪犯亦可能使用各種不同的藏密技術傳遞不法訊息，危害國土與社會安全。根據美國中央情報局及相關單位的調查，研判策劃 911 恐怖攻擊事件的蓋達組織，極可能利用藏密技術來傳送情報(Sieberg, 2001)；另外德國當局在其所破獲的恐怖分子嫌犯的偽裝色情檔案中，也發現了國際恐怖組織所隱藏的上百份秘密文件(陳文和，民 101 年)。除此之外，電腦罪犯更可能透過藏密技術，隱藏自己的相關犯罪行為與躲避相關單位之監控與查緝。以上例子不但說明藏密技術不法運用的可行性，更突顯出發展藏密分析技術(Steganalysis)之重要性與急迫性。

藏密分析技術依偵測對象，可概分為特定藏密分析(Specific Steganalysis)及通用藏密分析(Universal/Blind Steganalysis)等 2 大類。特定藏密分析技術係指針對特定的藏密演算法所發展的分析技術，經分析藏密演算法對載體影像所產生的變化後，以此變化作為特徵，對測試影像進行偵測，以確認其是否藏有秘密訊息。通用藏密分析技術綜整不同藏密演算法對藏密影像所產生的統計數據，建立特徵集合，再運用分類器、及統計學分類工具等智慧型運算技術進行特徵比對，來判斷載體影像中是否藏有秘密訊息，因此可偵測較多類之藏密技術，但偵測正確率沒有特定藏密分析技術來得高。

本論文旨在針對 Kim PEHS 藏密技術，提出有效的特徵，以有效偵測 Kim 所提出的 PEHS 藏密技術，其餘各節安排如下：第貳節探討基於預測差值直方圖位移之藏密技術、特定與通用藏密分析技術，並概述倒傳遞類神經網路；第參節說明本文提出的基於預測差值直方圖特徵之資訊隱藏分析技術；第肆節為實驗結果；最後一節為本文之結論。

貳、文獻探討

一、基於預測差值直方圖位移之藏密技術

Hong 等學者(2009) 提出基於 PEHS 的藏密技術，簡稱 Hong 藏密法，其嵌入程序如下：

步驟一：首先產生一個空的預測矩陣(I')，其大小與載體影像相同(I)，並將載體影像第 1 列與第 1 行的像素值複製到空的預測矩陣。

步驟二：自第 2 列及第 2 行開始依序掃瞄預測矩陣中的像素值，並使用 MED (Median Edge Prediction)預測器來產生載體影像的預測影像($\hat{I}'$)，即利用左上、上方及左方等 3 個相鄰點的像素值預測右下方位置的像素值(如圖 2所示)，如式(1)所示：

$$\hat{I}'_{i,j}=\begin{cases}\min\left(I'_{i,j-1},I'_{i-1,j}\right) & \text{if } I'_{i-1,j-1}\geq\max\left(I'_{i,j-1},I'_{i-1,j}\right)\\ \max\left(I'_{i,j-1},I'_{i-1,j}\right) & \text{if } I'_{i-1,j-1}\leq\min\left(I'_{i,j-1},I'_{i-1,j}\right)\\ I'_{i,j-1}+I'_{i-1,j}-I'_{i-1,j-1} & \text{otherwise}\end{cases},\quad(1)$$

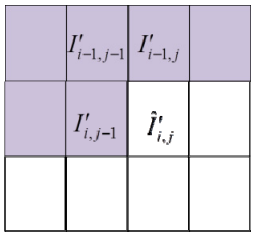

圖2　MED 預測

步驟三：利用式(2)計算載體影像與預測影像像素值的預測差值 e，

$$e=I_{i,j}-\hat{I}'_{i,j}。\quad(2)$$

步驟四：假如 e = 0 或–1，至步驟五進行藏密；否則，至步驟六進行像素值位移。

步驟五：依下列規則進行藏密：

(1)假如要嵌入的位元為‘0’，則 e 保持不變。

(2)假如要嵌入的位元為‘1’且 e = 0，將 e 的值修改為 $e+1$。

(3)假如要嵌入的位元為‘1’且 $e=-1$，將 e 的值修改為 $e-1$。

步驟六：假如 $e>0$，則將 e 的值修改為 $e+1$；假如 $e<-1$，則將 e 的值修改為 $e-1$。

步驟七：依式(3)產生藏密影像：

$$I'_{i,j}=\hat{I}'_{i,j}+e。\quad(3)$$

由於 PE 矩陣的值具有大量為 0 及–1 的特性，可產生較高的峰值點統計量，再加上利用 2 組峰值點進行位移藏密，此方法與基於 HS 的藏密技術相比較，其藏密量較高。

Tsai 等學者(2009)提出結合區塊分割之 PEHS 的藏密技術，簡稱 Tsai 藏密法，其嵌入程序如下：

步驟一：將影像分割成大小一致（3×3，4×4 或 5×5）的不重疊區塊，圖 3為不重疊的 3×3 像素區塊示意圖。

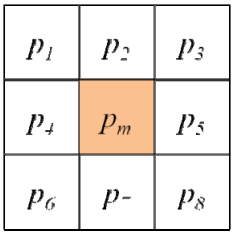

圖3　3×3 像素區塊示意圖

步驟二：利用式(4)，將位於區塊中間的像素值減去周圍的像素值以得到像素值的預測差值(PE)。

$$e_i=p_m-p_i，\quad(4)$$

其中，e_i 為預測差值，p_m 為 3×3 像素區塊的中間值，p_i 為相鄰像素值，而 i 的範圍為[1,8]。

步驟三：分別針對每一個 3×3 像素區塊進行 PE 的直方圖統計。

步驟四：分別從正與負的 PE 值域中各取 1 組峰值點，進行 HS 藏密。

Kim 等學者(2009)提出結合基於取樣影像之 PEHS 的藏密技術，簡稱 Kim PEHS 藏密法。其先利用式(5)將影像分割成 4 張子影像：

$$\begin{cases}s_1(i,j)=I(2i-1,2j-1)\\ s_2(i,j)=I(2i-1,2j)\\ s_3(i,j)=I(2i,2j-1)\\ s_4(i,j)=I(2i,2j)\end{cases},\quad(5)$$

其中 I 為載體影像，s_1、s_2、s_3 及 s_4 分別為分割後的 4 張子影像，(i,j)代表影像中列與行的索引值。再以其中一張子影像作為參考影像，計算其與其他子影像相對像素值之差值。以 s_1 為例，其與其它子影像之間的差值之計算方式如式(6)所示：

$$\begin{cases}d_2(i,j)=s_2(i,j)-s_1(i,j)\\ d_3(i,j)=s_3(i,j)-s_1(i,j)\\ d_4(i,j)=s_4(i,j)-s_1(i,j)\end{cases},\quad(6)$$

其中 d_2、d_3 及 d_4 分別為子影像 s_1 與參考影像 s_2、s_3 及 s_4 之間像素值的差值，其範圍為[-255, 255]。

Kim PEHS 藏密法屬於多層級的直方圖位移法，可依欲嵌入密文的數量決定直方圖位移的範圍，其利用 Embedding Level (*EL*)技術，使 PE 可以有較多次的修改，以獲取更大的藏密量。假設 *EL* 等於 2，可利用式(7)進行直方圖位移如下：

$$d'_k(i,j)=\begin{cases} d_k(i,j)+EL & \text{if } d_k(i,j)>EL \\ d_k(i,j)-EL & \text{if } d_k(i,j)<-EL \\ d_k(i,j) & \text{otherwise} \end{cases}, \quad (7)$$

其中 d'_k 為位移後的預測差值，$k\in\{2,3,4\}$。再利用式(8)進行藏密，如下：

$$d''_k(i,j)=\begin{cases} d'_k(i,j)+EL+w & \text{if } d'_k(i,j)=EL \\ d'_k(i,j)-EL-w & \text{if } d'_k(i,j)=-EL \end{cases}, \quad (8)$$

其中 w 為欲嵌入的秘密訊息，d''_k 為已嵌入秘密訊息 w 的預測差值。再利用式(9)將參考影像與已嵌入秘密訊息 w 的預測差值相結合，得到已嵌入秘密訊息的子影像如下：

$$\begin{cases} s'_2(i,j)=s_1(i,j)+d''_2(i,j) \\ s'_3(i,j)=s_1(i,j)+d''_3(i,j) \\ s'_4(i,j)=s_1(i,j)+d''_4(i,j) \end{cases}, \quad (9)$$

最後，將參考影像 s_1 與已嵌入秘密訊息的子影像 s'_2、s'_3 及 s'_4 相結合後，即得到藏密影像。

二、特定與通用藏密分析技術

Bender(1996)等學者所提的 LSB 藏密法，將載體影像像素的最不重要位元以秘密訊息來取代。若欲嵌入載體影像的秘密訊息二進位值之串流呈現均勻分佈(Uniform Distribution)，則此秘密訊息出現 0 或 1 的機率則幾乎相同，而載體影像像素的最不重要位元出現 0 或 1 的機率未必會相同。若將載體影像中每一像素的最不重要位元以秘密訊息取代，除了造成藏密影像最低位元統計量的改變外，也將造成藏密影像相鄰像素值之間關係的改變，RS(Fridrich et al., 2001)、x^2(Westfeld et al., 1999)與 SPA(Dumitrescu et al., 2003)等統計式的偵密技術即透過此現象來分析影像是否隱含秘密訊息。另外，Ker 學者(2007)則提出了結構化的統計分析偵密技術，可有效偵測使用二個 LSBs 的藏密。

大部分特定藏密分析技術均根據藏密影像的一階統計量來進行分析，如直方圖統計變化量或群組分佈變化量等。由於不同的藏密技術對藏密影像所造成的統計量的改變並不相同，若以特定藏密分析技術的觀點來看，針對不同藏密演算法就會產生一個相對應的藏密分析技術。若可使用一組共通的統計特徵即可偵測不同的藏密演算法，且偵測結果位於一定水準之上，在實務應用上更具效益。基於上述，Lyu 等學者(2006)年提出一個基於高階影像統計量的藏密分析技術，實驗結果顯示，針對一般之藏密方法，有不錯的偵測結果。

Geetha 等學者(2009)提出與影像內容獨立的影像品質矩陣(Content Independent Image Quality Metrics, CIIQM)作為特徵，並結合基因演算法與 X-means 分類技術，提出通用藏密分析技術。實驗結果顯示，針對 JPEG 影像之相關藏密法，其平均偵測正確率可達 86.54%。Gul 等學者(2010)提出通用藏密分析技術，其先將影像切割成大小為 $W\times W$ ($W\in$ [3,27])的子區塊，以每個特定 W 區塊的奇異值分解(Singular Value Decomposition, SVD)值作為特徵，並結合維納濾波(Wiener Filter)與支援向量機(Support Vector Machine, SVM)進行偵測。實驗結果顯示，針對空間域影像之相關藏密法(LSB、LSB Matching 與 Steghide)，其偵測正確率最高可達 91.18%，而針對頻率域影像之相關藏密法(F5、JP Hide&Seek、Outguess、MB1、MB2 及 PQ)，其偵測正確率最高可達 72.97%。Pevny 等學者(2010)利用統計學中的馬可夫鏈(Markov Chain)的概念，提出通用藏密分析技術。實驗結果顯示，針對空間域藏密法與 JPEG 相關的藏密法有良好的偵測成效。

三、倒傳遞類神經網路

McCulloch 及 Pitts 二位學者(1943)共同提出結合神經生理學及邏輯數學的神經元模型，開始類神經網路之研究。而更進一步的發展則是由 Rosenblatt(1958)提出的感知器(Perceptron)架構，其成功模擬生物的感知與學習能力，引發了學者對類神經網路的研究風潮。後續因應不同的問題類型，產生出許多不同型態的類神經網路。

Rumelhart 等學者(1986)提出倒傳遞類神經網路(Back-Propagating Neural Network, BPNN)，其適用於處理複雜的高度非線性函數合成問題。BPNN 是由多層的神經元結構所構成，其中第一層接收輸入變數，稱為輸入層(Input Layer)，中間的神經元稱為隱藏層(Hidden Layer)，而最後產生預測結果的神經元則稱之為輸出層(Output Layer)。隱藏層的功能主要是增加類神經網路的複雜性，以便模擬複雜的非線性關係，BPNN 的架構如圖 4所示。

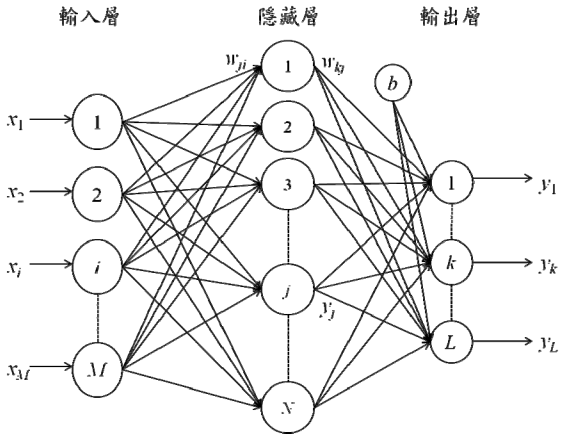

圖4　倒傳遞類神經網路架構圖

BPNN 架構的輸出值為

$$y_k = f\left(\sum_{j=1}^{N} w_{kj} y_j - b_k\right) , \tag{10}$$

其中 y_k 為輸出層神經元 k 的輸出值；y_j 為隱藏層神經元 j 的輸出值；$f(\bullet)$為非線性轉移函數；w_{kj} 為隱藏層神經元 j 與輸出層神經元 k 之間的權重值；b_k 為輸出層神經元 k 的偏權值。

由於 BPNN 屬於監督式學習網路，針對每筆輸入資料，都有其相對應的目標輸出值(d_k)可供比對。故在每次迭代的過程中，均會計算網路輸出值(y_k)與目標輸出值(d_k)之間的均方差(Mean Square Error, MSE)

$$e(t) = \frac{1}{2}\sum_{k}\left(d_k(t) - y_k(t)\right)^2 , \tag{11}$$

BPNN 調整網路連結權重的學習法則是利用最陡坡降法找尋最小的瞬時目標函數值。假設目標函數值為目標輸出值與 BPN 網路輸出值的誤差期望值（即 $e(t)$），則對目標函數微分後得

$$\Delta w_{ji} = \eta \delta_j^n y_i^{n-1} , \tag{12}$$

其中

$$\delta_j^n = -\frac{\partial e}{\partial net_j^n} 。 \tag{13}$$

在每一次學習的過程中，可利用式(12)進行輸入層與隱藏層間或隱藏層與輸出層間的神經元之權重值與偏權值的調整，直到網路輸出值達到容忍誤差值標準或學習迭代的次數達到設定的最大值。

參、基於預測差值直方圖特徵之資訊隱藏分析技術

本節提出可偵測 Kim 等學者提出的基於 PEHS 藏密法之偵測技術(以下稱為本偵測技術)，其主要是針對藏密後之藏密影像之 PEH 特徵為基礎，結合 BPNN 分類器進行藏密偵測。本節第一部分分析 Kim 等學者提出的基於 PEHS 藏密法在嵌密過程中所產生之 PEH，第二部分闡述本偵測技術針對 Kim PEHS 藏密方法所提出之特徵值，第三部分則說明本偵測技術之偵測流程。

一、Kim PEHS 藏密法之分析

為了找出適用於 Kim PEHS 藏密法的藏密特徵，本論文利用區塊預測器將載體與藏密影像區分為 4 個區塊，選定 1 個基準區塊後，其餘 3 個區塊像素值減去基準區塊像素值，可得預測差值(PE)。再比對

載體與藏密影像的預測差值直方圖(PEH)，找出其相異特徵，以作為有效區分載體與藏密影像之基礎，其特徵比對流程如圖 5所示。

依圖 5的流程，可得載體影像與 Kim PEHS 藏密法所產生的藏密影像之 PEH 的對比圖，如圖 6所示。從圖 6可明顯觀察出載體影像與 Kim PEHS 藏密法所產生的藏密影像之 PEH 有明顯的差異，可藉此建構出有用的特徵。

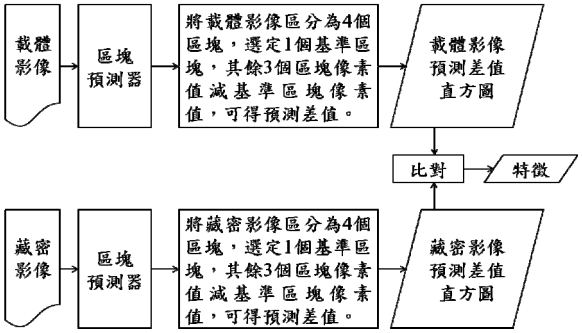

圖5　載體與藏密影像特徵比對流程圖

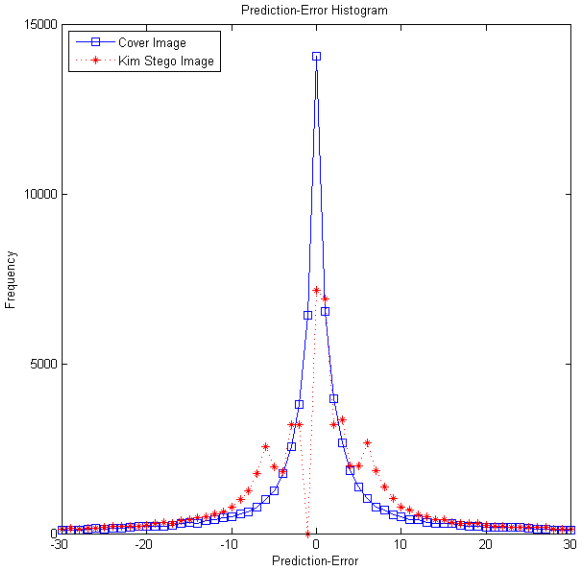

圖6　載體與藏密影像之 PEH 對比圖

二、本偵測技術使用之特徵值

由於秘密訊息嵌入的緣故，載體影像的 PEH 產生顯著改變，其一為載體影像的 PEH 的數值劇烈地下降；其二為由於 Kim 藏密影像不同區塊 PE 位移所造成的變化，使得 Kim 藏密影像各區塊 PE 為-1 的數值為 0，與載體影像各區塊 PE 為-1 的數值相比，這是非常明顯的特徵，如圖 7與圖 8所示。PE 值為-3 至 3 的 PEH 示意圖如圖 9所示，其中 H_{-3}、H_{-2}、H_{-1}、H_0、H_1、H_2 及 H_3 分別代表 PE 值為-3、-2、-1、0、1、2 及 3 的統計量，本論文以 PEH 之 PE 比率關係作為藏密之 3 項特徵，整理如表 1。

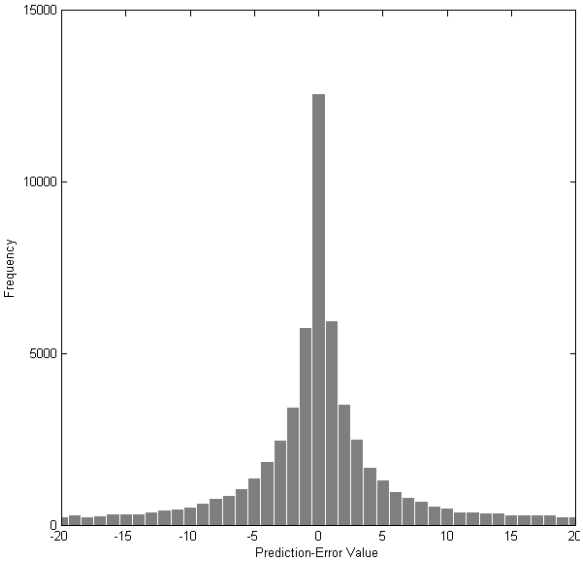

圖7　載體影像的 PEH

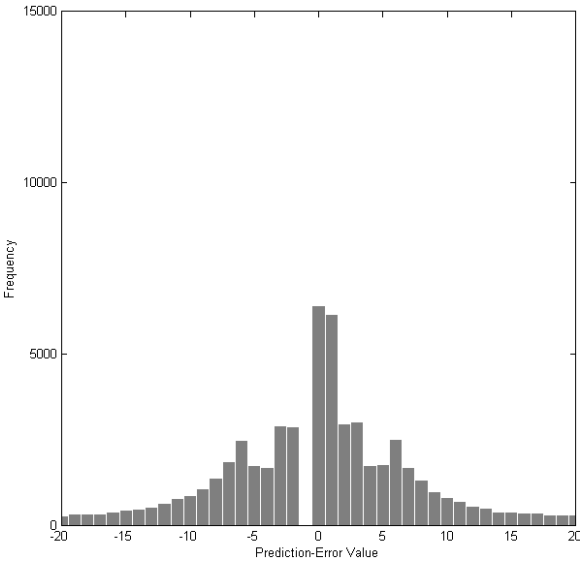

圖8　藏密影像的 PEH

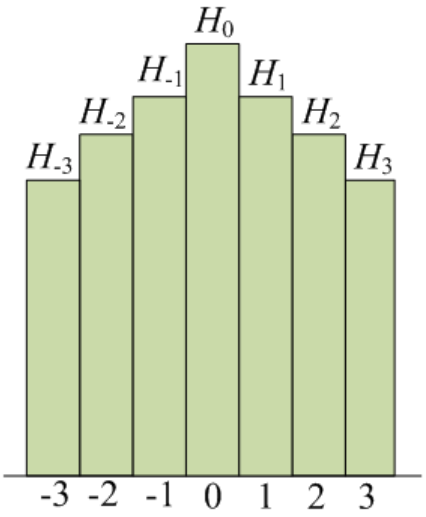

圖9　PE 值範圍為-3 至 3 的 PEH 示意圖

表1　本偵測技術所提之特徵值

特徵值符號	特徵值計算
F_1	$\frac{H_0}{(H_{-1}+1)}$
F_2	$\frac{H_{-6}}{H_{-2}}$
F_3	$\frac{H_6}{H_2}$

為了說明 F_1 至 F_3 特徵值的有效性，本論文針對實驗所使用的 NRCS 影像資料庫中 2724 張載體影像及相對應的 Kim PEHS 藏密影像進行特徵值擷取，其結果如圖 10至圖 12所示，其橫軸為測試影像，而縱軸則代表影像之 F_1 至 F_3 的特徵值。

圖 10至圖 12之結果顯示，載體影像的特徵值 F_1 至 F_3 呈現亂數的分佈情形；但對於藏密影像而言，因 Kim PEHS 藏密技術對於像素值的調整，致使特徵值 F_1 至 F_3 的值呈現群聚現象，故載體影像與藏密影像 F_1 至 F_3 特徵值的分佈呈現兩明顯分隔的群組。

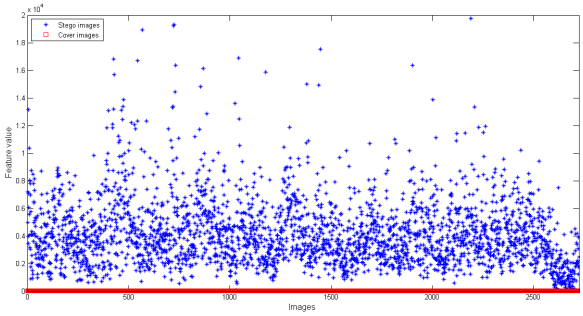

圖10　載體影像與 Kim PEHS 藏密影像之 F_1 特徵值分佈圖

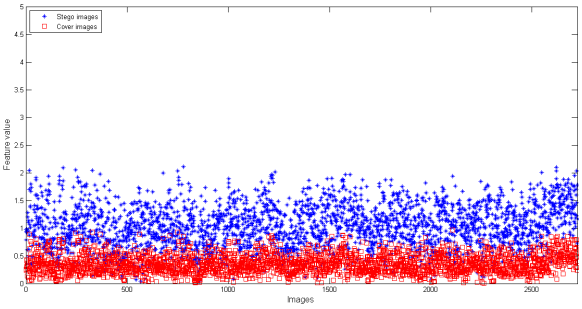

圖11　載體影像與 Kim PEHS 藏密影像之 F_2 特徵值分佈圖

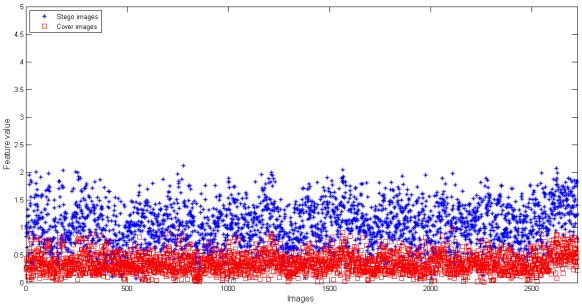

圖12　載體影像與 Kim PEHS 藏密影像之 F_3 特徵值分佈圖

圖 13為本論文使用區塊預測器將測試影像區分為 3 組不同的差值區塊 (d_2-d_4)，這三組不同的差值區塊分別利用表 1。產生 3 個特徵，故本論文共使用 9 個特徵供本偵測技術之偵測流程使用。

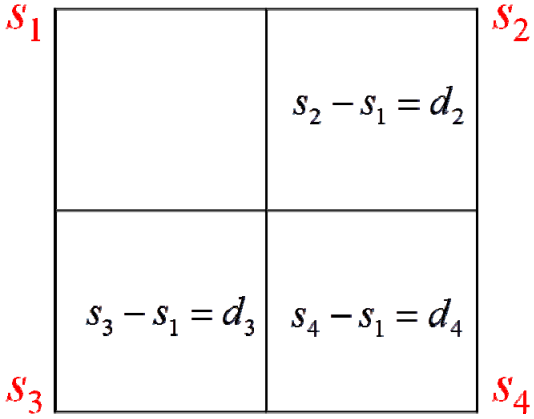

圖13　本論文使用的 3 組差值區塊

三、本偵測技術之偵測流程

本偵測技術主要在結合本論文所使用的 9 個特徵值與 BPNN 分類器的特性，以偵測待測影像是否為使用 Kim PEHS 藏密技術之藏密影像，其偵測流程如圖 14 所示。本偵測技術概分為兩大階段，一為分類模型產生階段，一為影像偵測階段。

在分類模型產生階段，本研究利用未藏密的載體影像與經 Kim PEHS 藏密演算法藏密後的藏密影像作為訓練樣本，並根據本論文所提的 9 項特徵值對所有樣本影像進行特徵值擷取，完成訓練樣本特徵值集合。另外，本論文分別設定藏密影像之分類標籤為 1 而載體影像分類標籤為 2，並將分類標籤與訓練樣本特徵值輸入 BPNN 分類器進行訓練與學習，使 BPNN 分類器能學習到從藏密影像所擷取的特徵集合對應於分類標籤 1；從載體影像所擷取的特徵集合對應於分類標籤 2，以產生訓練模型(Trained Model)。

在影像偵測階段，我們亦將受測影像根據本論文所提的 9 項特徵值進行特徵值擷取，並輸入至 BPNN 分類器進行藏密偵測。BPNN 分類器依據已完成訓練的模型對受測影像進行分類，根據先前 BPNN 分類模型產生階段所學習到藏密影像與載體影像之特徵集合的對應方式，輸出分類結果。若輸出之標籤為 1，表示為使用 Kim PEHS 藏密法所產生的藏密影像；若輸出之標籤為 2，表示為載體影像。

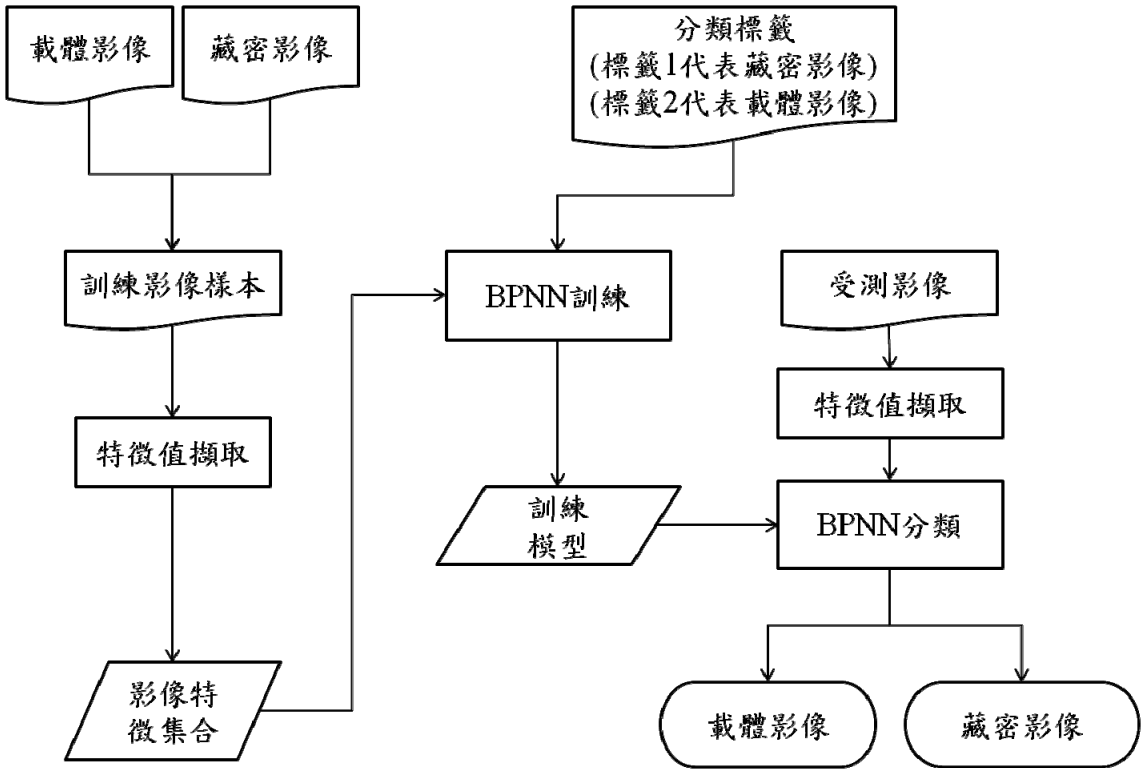

圖14　本偵測技術之偵測流程圖

肆、實驗結果

本實驗採用 NRCS 自然影像資料庫 8 位元 512*512 大小之 2724 張灰階影像與 UCID 影像資料庫 8 位元 512*384 及 384*512 的 1338 張灰階影像做為原始影像(範例影像如圖 15及圖 16所示) 及使用 BPNN 分類器。不失其一般性，本實驗以 MATLAB 內建之亂數產生器產生實驗所需之均勻分佈二位元數值序列，模擬加密後之秘密訊息。為驗證本偵測技術之效果，本研究以 MATLAB 實現 Kim PEHS 藏密技術、特徵值萃取程式，及實現 Zhang(2007) 等學者所提的 ALE 偵測技術與 Pevný 等學者(2010)所提 SPAM 偵測技術。相關實驗步驟及結果分述於以下小節。

圖15　NRCS 資料庫之灰階範例影像

圖16　UCID 資料庫之灰階範例影像

一、實驗步驟

步驟一：在NRCS的2724張資料庫影像中，隨機挑選1362張影像，使用Kim PEHS進行藏密，得到Kim PEHS藏密影像1362張。

步驟二：利用特徵值萃取程式，分別針對未藏密的 1362 張載體影像及已藏密的1362張Kim PEHS藏密影像進行特徵值萃取。

步驟三：將步驟二所得之2724張影像特徵值及相對應的分類標籤輸入BPNN分類器進行分類訓練，產生分類模型。

步驟四：將資料庫中扣除訓練用影像外的1362張自然影像作為載體影像，進行Kim PEHS藏密，產生1362張藏密影像。

步驟五：將步驟三所得之分類模型及步驟四所得之2724張影像(1362張載體影像和Kim PEHS藏密影像)分別輸入BPNN分類器進行分類，並記錄其分類結果。

步驟六：重複進行十次步驟四及五，求得實驗結果之平均值。

而針對 UCID 影像資料庫的實驗步驟與上述針對 NRCS 影像資料庫的方式相同，亦先針對 1338 張 UCID 資料庫影像中，隨機挑選 669 張影像，使用 Kim PEHS 進行藏密，得到 Kim PEHS 藏密影像 669 張，而後續步驟 2 至 6，則與前述步驟 2 至 6 相同。

二、實驗結果

為依上小節之實驗步驟所得實驗結果，其中 *AC* 值為偵測正確率，定義如下：

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \text{。} \quad (14)$$

其中 *TP* 為正確判別藏密影像之張數、*FP* 為錯誤判別藏密影像之張數、*TN* 為正確判別未藏密影像之張數、*FN* 為錯誤判別未藏密影像之張數。

本偵測技術的實驗結果如表 2及表 3所示，由表 2及表 3可看出本技術其偵測正確率平均可達 99.9%與 99.6%，可證明本偵測技術可有效偵測使用不同影像大小之灰階影像的 Kim PEHS 藏密技術。

表2　本偵測技術針對 NRCS 影像資料庫之偵測結果

項次	正確分類之影像張數				*AC* 值
	TP	*FN*	*TN*	*FP*	
1	1359	3	1362	0	99.9%
2	1359	3	1362	0	99.9%
3	1359	3	1362	0	99.9%
4	1358	4	1362	0	99.9%
5	1358	4	1362	0	99.9%
6	1359	3	1362	0	99.9%
7	1362	0	1360	2	99.9%
8	1359	3	1362	0	99.9%
9	1357	5	1362	0	99.8%
10	1359	3	1362	0	99.9%
平均(%)	1358.9	3.1	1361.8	0.2	99.9%

表3　本偵測技術針對 UCID 影像資料庫之偵測結果

項次	正確分類之影像張數				*AC* 值
	TP	*FN*	*TN*	*FP*	
1	664	5	665	4	99.3%
2	662	7	669	0	99.5%
3	668	1	666	3	99.7%
4	667	2	667	2	99.7%
5	665	4	667	2	99.6%
6	667	2	668	1	99.8%
7	663	6	669	0	99.6%
8	665	4	669	0	99.7%
9	666	3	668	1	99.7%
10	668	1	663	6	99.5%
平均(%)	665.5	3.5	667.1	1.9	99.6%

表4　本偵測技術與 ALE 及 SPAM 偵測正確率(%)之比較

藏密技術	偵測技術		
	本偵測技術	ALE	SPAM
Kim PEHS	99.9%	69.7%	70.9%

為顯示本偵測技術之偵測績效，本分析技術與 Zhang 等學者(2007)所提的偵測技術(ALE) 及 Pevný 等學者(2010)所提的空間域通用偵測技術(SPAM)進行比較，比較結果如表 4 所示。其可明顯看出，本分析技術偵測 Kim PEHS 藏密法之偵測正確率均優於 ALE 與 SPAM。此結果說明，本分析技術使用的特徵數量雖然只有 9 個，少於 Zhang 等學者所提出的 10 個特徵與 Pevný 等學者提出的 686 個特徵，但偵測正確率均遠優於 SPAM，且所耗費的計算資源與時間亦低於 ALE 及 SPAM，可證本分析技術之優越性。

伍、結論

藏密技術之濫用可能危害國土與社會安全，也突顯出發展藏密分析技術重要性與急迫性。本論文旨在針對 Kim PEHS 藏密技術，提出有效的特徵，以有效偵測 Kim PEHS 藏密技術。

本偵測技術利用區塊預測器將載體與藏密影像區分為 4 個區塊，選定 1 個基準區塊後，其餘 3 個區塊像素值減去基準區塊像素值，可得預測差值(PE)。再比對載體與藏密影像的預測差值直方圖(PEH)，找出其相異特徵，以作為有效區分載體與藏密影像之基礎。本論文之偵測流程區分為訓練模型產生及影像偵測兩個階段。在訓練模型產生階段，先輸入已標記的載體影像與藏密影像樣本至 BPNN 分類器，進行學習與訓練，產生訓練模型。在影像偵測階段，則利用已訓練的模型，對輸入之受測影像進行偵測分類。

為驗證本偵測技術所使用藏密特徵之有效性，本研究使用 MATLAB 實現 Kim PEHS 藏密技術之模擬程式及特徵值擷取程式，並結合了 BPNN 分類器及 NRCS 影像資料庫以進行實驗。實驗結果證明，本研究所提之 9 項 Kim PEHS 藏密特徵於偵測 Kim PEHS 藏密法時，可達 99.9%的偵測正確率。實驗結果同時證明，本偵測技術對 Kim PEHS 藏密技術之偵測效果均優於 ALE 及 SPAM 通用型之偵測技術。

近期國際上發生多起恐怖攻擊事件，在維護國家與國防安全的趨勢下，須設法防止不法人士於網路上利用資訊隱藏方法來傳送訊息。而本偵測技術的研究成果，將可應用於網際網路上偵測是否有可疑資訊，進而防患於未然，保護國家及國防安全。

本偵測技術之研究限制有以下二點：

一、本論文所提出的 Kim PEHS 藏密分析技術，是假設 Kim PEHS 藏密技術使用的載體影像為灰階影像。若 Kim PEHS 藏密技術使用的載體影像為彩色影像時，本研究提出的藏密分析技術是否能適用，這是本研究值得進一步探究的議題。

二、本論文基於載體影像與 Kim PEHS 藏密法之藏密影像的 PEH 進行比較，設計有效之特徵，以作為有效區分載體與 Kim PEHS 藏密影像之基礎。但若偵測非採用 PEHS 藏密法所產生的藏密影像，因其藏密影像之 PEH 不會產生 Kim PEHS 藏密法的藏密特徵，故本偵測技術無法進行有效之偵測。

參考文獻

陳文和(民 101 年 5 月 2 日)。*暗藏基地情報 A 片變恐怖片*。**中時電子報**。取自 http://news.chinatimes.com/world/110504/112012050200143.html.

Bender, W., Gruhl, D., Morimoto, N., and Lu, A., 1996. Techniques for Data Hiding, *IBM Systems Journal*, 35(3-4), 313-336.

Cacciaguerra, S. and Ferretti, S., 2000. *Data Hiding: Steganography and Copyright Marking*, Department of Computer Science, University of Bologna, Italy.

Dumitrescu, S., Wu, X., and Wang, Z., 2003. Detection of LSB Steganography via Sample Pair Analysis, *IEEE Transactions on Signal Processing*, 51(7), 906-910.

Fallahpour, M. and Sedaaghi, M. H., 2007. High Capacity Lossless Data Hiding Based on Histogram Modification, *IEICE Electronics Express*, 4(7), 205-210.

Feng, J. B., Lin, I. C., Tsai, C. S., and Chu, Y. P., 2006. Reversible Watermarking: Current Status and Key Issues,

International Journal of Network Security, 2(3), 161-171.

Fridrich, J., Goljan, M., and Rui, D., 2001. Detecting LSB Steganography in Color, and Gray-scale Images, *Magazine of IEEE Multimedia Special Issue on Security*, 4(4), 22-28.

Geetha, S., Sindhu, S., and Kamaraj, N., 2009. Blind Image Steganalysis Based on Content Independent Statistical Measures Maximizing the Specificity and Sensitivity of the System, *Computers & Security*, 28(7), 683-697.

Gul, G. and Kurugollu, F., 2010. SVD Based Universal Spatial Domain Image Steganalysis, *IEEE Transactions on Information Forensics and Security*, 5(2), 349-353.

Hong, W., Chen, T. S., and Shiu, C. W., 2009. Reversible Data Hiding for High Quality Images Using Modification of Prediction Errors, *Journal of Systems and Software*, 82(11), 1833-1842.

Hwang, J. H., Kim, J. W., and Choi, J. U., 2006. A Reversible Watermarking Based on Histogram Shifting, *Lecture Notes in Computer Science*, 4283, 348-361.

Ker, A. D., 2007. Steganalysis of Embedding in Two Least-Significant Bits, *IEEE Transactions on Information Forensics and Security*, 2(1), 46-54.

Kim, K. S., Lee, M. J., Lee, H. Y., and Lee, H. K., 2009. Reversible Data Hiding Exploiting Spatial Correlation between Sub-sampled Images, *Pattern Recognition*, 42(11), 3083-3096.

Lyu, S. and Farid, H., 2006. Steganalysis Using Higher-order Image Statistics, *IEEE Transactions on Information Forensics and Security*, 1(1), 111-119.

McCulloch, W. and Pitts, W., 1943. A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biology*, 5(4), 115-133.

Mielikeainen, J., 2006. LSB Matching Revisited, *IEEE Signal Processing Letters*, 13(5), 285-287.

Ni, Z., Shi, Y. Q., Ansari, N., and Su, W., 2006. Reversible Data Hiding, *IEEE Transactions on Circuits and Systems for Video Technology*, 16(3), 354-362.

Paar, C. and Pelel, J., 2010. *Understanding Cryptography*, Springer-Verlag, London.

Petitcolas, F. A. P., Anderson, R. J., and Kuhn, M. G., 1999. Information Hiding-A Survey, *Proceedings of the IEEE*, 87(7), 1062-1078.

Pevný, T., Bas, P., and Fridrich, J., 2010. Steganalysis by Subtractive Pixel Adjacency Matrix, *IEEE Transactions on Information Forensics and Security*, 5(2), 215-224.

Rabah, K., 2004. Steganography-The Art of Hiding Data, *Information Technology Journal*, 3(3), 245-269.

Rosenblatt, F., 1958. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Review*, 65(6), 386-408.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986. Learning Representations by Back-Propagating Errors, *Nature*, 323(9), 533-536.

Sieberg D., 2001. *Bin Laden exploits technology to suit his needs*. From http://edition.cnn.com/2001/US/09/20/inv.terrorist.search/ (retrieved on March 18, 2015).

Tsai, P., Hu, Y. C., and Yeh, H. L., 2009. Reversible Image Hiding Scheme Using Predictive Coding and Histogram Shifting, *Signal Processing*, 89(6), 1129-1143.

Tseng, H. W. and Hsieh, C. P., 2008. Reversible Data Hiding Based on Image Histogram Modification, *Imaging Science Journal*, 56(5), 271-278.

Westfeld, A. and Pfitzmann, A., 1999. Attacks on Steganographic Systems, *Proceedings of the Third International Workshop on Information Hiding*, Dresden, Germany, 61-75.

Zhang, J., Cox, I.-J., Doerr, G., 2007. Steganalysis for LSB Matching in Images with High-frequency Noise, *IEEE 9th Workshop on Multimedia Signal Processing*, 385-388.