二元文件竄改位置偵測與還原及其延伸應用

王中全[1] 黃詩迪[2] 蕭瑞霈[3]

[1] 中州科技大學資訊工程系 ccwang@dragon.ccut.edu.tw

[2] 中州科技大學工程研究所 chbits0000@gmail.com

[3] 中州科技大學資訊工程系 nba781111@hotmail.com

摘 要

資訊隱藏著重在藏密於多媒體且不被察覺性的條件下，過去雖有許多文件竄改偵測與還原機制設計於灰階及彩色圖像上，但日常生活中有許多重要文件是黑白文字文件，卻少有設計於二元文件竄改偵測與內容還原機制，無法對所有重要文件提供有效的保護。有鑑於此，我們提出一個新的二元文件偵測竄改位置及內容還原機制，並延伸此法到灰階及彩色圖像上，達成重要文件的保護機制。我們為加大藏量足以藏文件本身內容數份於於其中，採用2X2小區塊來增加藏密區塊數量，每一小區塊藏量採變動長度藏密使藏量極大化。在減少變動量上，藏入區塊採2互補區塊對應同一密文藏密，保證任何2X2藏密區塊變動不超過2位元。在偵測文件竄改位置上採每3個藏密區塊即加入1個檢核區塊來達成有效偵測竄改位置，並藉由藏入多份內容於文件中，竄改文件後，內容得以多數決與檢核碼的方式達成文件內容復原的目標。因為黑白影像只可藏密於黑白邊緣，許多灰階及彩色文件的研究無法轉移到黑白影像，反之灰階影像則可依每一像素位元組中分別取出對應位元成為八張位元圖（bit plane），LSB-1，2，3（Least Significant Bits-1，2，3）方法即為對應到八張黑白位元圖中的最後三張，像素值（僅1，2，4）變動較小可任意藏入秘密，關鍵黑白位元圖如LSB-4像素值8，僅更動邊緣對原圖仍保有不易察覺性，此機制復原文件內文搭配LSB1~3灰階或彩色復原原圖機制，達成灰階或彩色重要文件竄改位置偵測與還原的雙重保護機制。

關鍵詞：竄改、偵測、變動長度、復原

通訊作者
姓名：王中全
E-mail：ccwang@dragon.ccut.edu.tw

壹、緒 論

科技快速發展，網路資訊流通頻繁，雖然帶給人們便利，相對地也衍生出安全性的問題，如欄截、竄改、破解…等各種不合法的網路犯罪日益嚴重，因此資訊安全議題越顯重要。資訊安全可概分為兩類：一、密碼學-著重文件內容的隱藏，如：DSA[8]、RSA[10]、…等，二、藏密學-著重秘密存在性的隱藏如：資訊隱藏[2, 6, 9,11, 12, 13,14]，數位浮水印[16]、視覺密秘分享[3]、…等，其中密碼學將內文轉換成亂碼，唯亂碼容易被察覺藏密而遭致攻擊，而強調不易察覺性的藏密學，密秘藏於多媒體文件（如圖片、音樂、影片、…等等）中，不易遭到攻擊，進一步將密碼學中的密文以藏密學的機制隱藏至多媒體中，可提供更周全的保護。二元影像擁有檔案小且傳輸快及列印成本低的優點，二元影像文件廣泛被使用，然而相較於灰階及彩色圖像，二元圖像藏密一直被認為是一大挑戰，在不易察覺的情況下，二元影像只能藏密於黑白邊緣，因此藏密量不大是最重要的原因，2006 年 Yang 等曾提出二元圖像的認證與竄改位置偵測[15]，但無法同時達到灰階及彩色圖像提出的竄改位置偵測及復原機制的兩大認證目標，歸納原因為過去的二元圖像藏密機制藏量小且視覺品質不佳，因此要解決此問題即需加大二元圖像藏量同時解決視覺品質，因此如何選取區塊型態增加數量及區塊內提高位元藏量和降低變動量是重要關鍵，在區塊型態方面，許多不同的設計以選取較大區塊如 3×3[9，13，15]，4×5[12]，4×4[3]，1×6[11]，1×4[6]，等大區塊只藏 1 位元，變動量與藏密量接近的結果，造成低藏量及高變動量，且邊緣長度大於 2 易產生毛邊及胡椒鹽現象。本篇論文提出滿足邊緣變動不易察覺特性的 2×2 小區塊為藏密區塊型態，在增加藏量上，小區塊選取除增加區塊數量，區塊藏密更以變動長度方式提高區塊藏秘位元，在視覺品質上，提出兩互補區塊對應同一藏秘方式，保證任何 2×2 區塊藏密變動絕不超過 2 位元大幅降低變動量，參考 parity check 機制提出 3 個藏密區塊即加入 1 個檢核區塊，達成二元文件竄改位置偵測目標，並藏入多份文件內容達成內容還原。而重要文件除二元圖像外，亦包含灰階或彩色圖像，因此我們的研究不僅應用在二元圖像，更可延伸於灰階及彩色圖像上，灰階圖像像素含 8 位元，使用 bit plane 的方法將灰階圖像分成 8 張二元圖像[1]；而彩色圖像像素含 3 組 8 位元，可分成 3 組 8 張二元圖像，取出 LSB1~4 的二元圖像，其中一般的藏密法多應用於 LSB1~3，而我們二元圖像因藏秘於邊緣，較不易查覺，配合我們的高藏量低變動量，因此可於 LSB4 的二元圖像上加入另一組不同的藏密機制來保護重要文件內容，與一般 LSB1~3 的灰階彩色復原原圖像素的方式相輔相成，因此可達成重要文件竄改位置偵測與還原的雙重保護機制。

貳、研究方法

一、 藏密

藏密部份可由文件圖像切割與藏密區塊轉碼與文件內文碼轉換與藏密兩部份說明

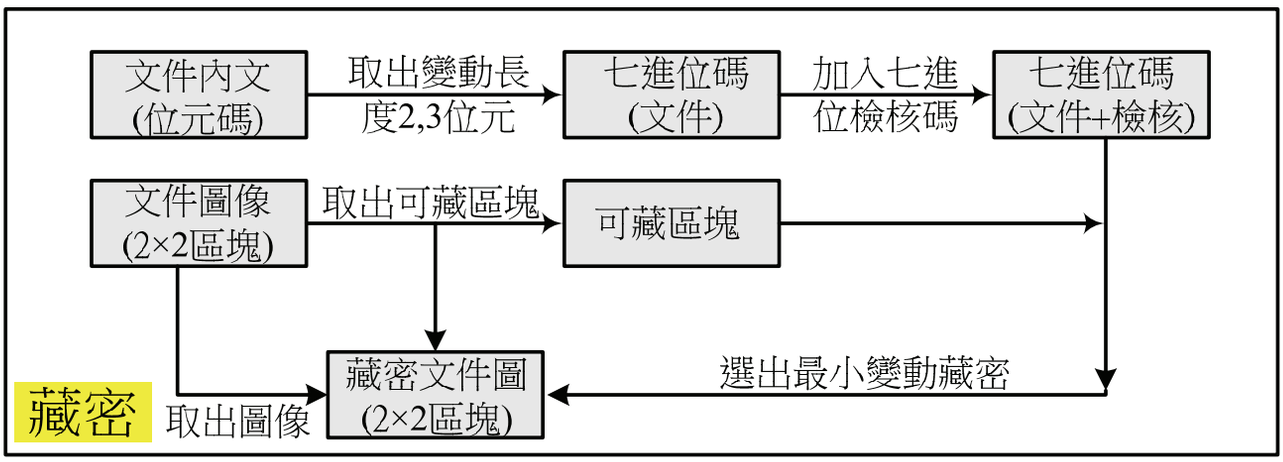


圖 1　藏密流程圖

（一）文件圖像

1、圖像切割與藏密

文件圖像 *I* 均分成四等份 I_1, I_2, I_3, I_4 如圖 2，並在四圖像中切割不交錯的 2×2 區塊，

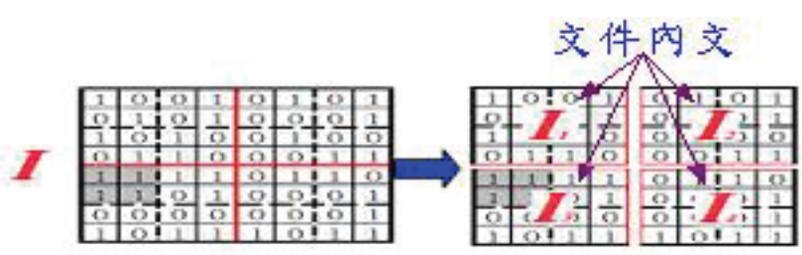


圖 2　原圖像均分四個掩護圖像

文件內文則重複存於四圖像 I_1, I_2, I_3, I_4 中，竄改後仍可藉由比對方式萃取出完整的文件內文。

2、2×2可藏密區塊轉碼

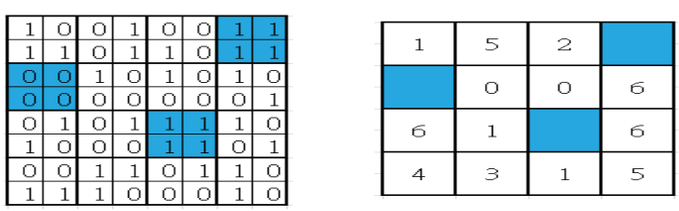
2×2 可藏密區塊就是非全黑或非全白的 14 種 2×2 區塊，分成 7 對互補區塊對，分別以七進位 *se* 0、1、2、3、4、5、6 來表示。

如表 1：簡單的分成兩類（1）非對稱互補區塊對（簡稱 NCB）-包含 1 黑 3 白區塊或 3 黑 1 白區塊（2）對稱互補區塊對（簡稱 SCB）-包含 2 黑 2 白區塊。

表 1　2×2 藏密互補區塊對與七進位數對應關係

2×2區塊可藏區塊	非對稱互補區塊對-NCB（1黑3白或3黑1白）				對稱互補區塊對-SCB（2黑2白）		
型態(*i*)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1 2 3 4 （位元順序）	(0111) (0111)	(0111) (0111)	(0111) (0111)	(0111) (0111)	(0111) (0111)	(0111) (0111)	(0111) (0111)
七進位(*se*)	**0**	**1**	**2**	**3**	**4**	**5**	**6**

對 NCB 區塊對而言，七進位 *se* = *i*-1，其中 *i* 為型態編號，代表位元順序中位元唯一不同於其他 3 位元像素，如 0001 和 1110 為型態 4 區塊對應到七進位值 3。對 SCB 區塊對而言，型態 5、6、7 分別對應七進位 *se*4、5、6，並以顏色叢集方向，水平（0011，1100）表 *se* 4，垂直（0101，1010）表 *se* 5，和斜向（0110，1001）表 *se* 6。例：一圖像 *H* 如圖 3（a）其中有 13 個 2×2 區塊可藏轉成七進位為 1520066164315 如圖 3（b）。



（a）　　（b）

圖 3　二元圖像 2×2 可藏區塊與七進位間轉換例

（二）　文件內文碼轉換與藏密

1、位元碼：重要文件內容可依中文內碼及英文 ASCII 碼轉成二進位文件碼儲存。

2、變動長度七進位碼：二進制文件內文以變動長度方式轉成七進位，如表 2，除"11"2 位元對應七進位"6"外，其餘採三位元一般對應,即“000”，“001”，…，“101”對應七進位“0”,“1”,…,“5”。可藏區塊位元藏量為 2.857 位元/區塊，一 2×2 區塊相當於 4 位元，可藏區塊位元藏密率 0.714（≈2.857/4）。

表 2　變動長度編碼（2、3 位元轉 7 進位碼）

單位	變動長度編碼							區塊位元藏量（位元/區塊）
2、3 位元 *a*	000	001	010	011	100	101	11	2.857
Septenary *se*	0	1	2	3	4	5	6	（≈（3×6+2）位元/7 區塊）

3、文件檢核碼：文件內文七進位碼依下列公式加入檢核碼，設連續每三個七進位數值 se_{4i}，se_{4i+1} 和 se_{4i+2} 加入一個檢核碼（se_{4i+3}），

$$se_{4i+3}=\begin{cases}(se_{4i}\times se_{4i+1}\times se_{4i+2}) \bmod 7 & \text{當 } se_{4i}\times se_{4i+1}\times se_{4i+2}\neq 0\\ (se_{4i}+se_{4i+1}+se_{4i+2}) \bmod 7 & \text{當 } se_{4i}\times se_{4i+1}\times se_{4i+2}=0\end{cases} \quad \text{(公式 1)}$$

例如藏入文件內文位元“011001100”，其中 011-se_0=3，001-se_1=1，100-se_2=4 則檢核碼 se_3=3×1×4 mod 7 = 5，若文件內文位元為“11000010”，則 11-se_0=6，000-se_1=0，010-se_2=2 則檢查碼 se_4=（6+0+2） mod 7 = 1 如圖 4。

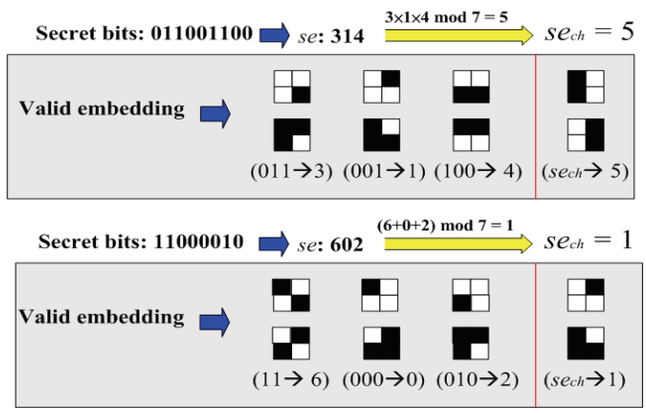


圖 4　檢核碼產生機制

為明晰此機制，圖 5 例中可藏區塊轉成七進位為”4652 4401 6054 4033”。其中 4652(4×6×5 mod 7=1 不是 2）與 4033（(4+0+3）mod 7=0 不是 2）是不合法的區塊組，達成區塊組竄改偵測的目標。

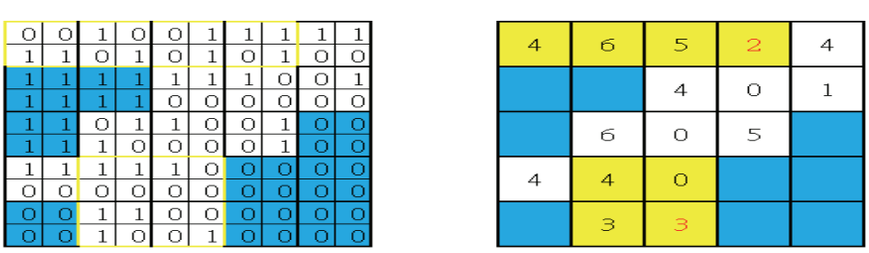


圖 5　區塊檢核實例

4、藏密

將文件內文重覆藏入 I_1、I_2、I_3 和 I_4 四個文件圖像，區塊像素最大對應法是由二對一的對應中找出最小的變動對應區塊。即兩種對應藏入區塊如表一，選出最小變動的藏入區塊，藏密機制如圖 6，函數 CB（se）表七進位 se 兩互補區塊如表 1，函數 Ω（被藏區塊，藏密取代區塊-CB（se））從 CB（se）中找出最少變動的區塊，例：$\Omega(0100,CB(5))=\Omega(0100,(0101,1010))=0101$

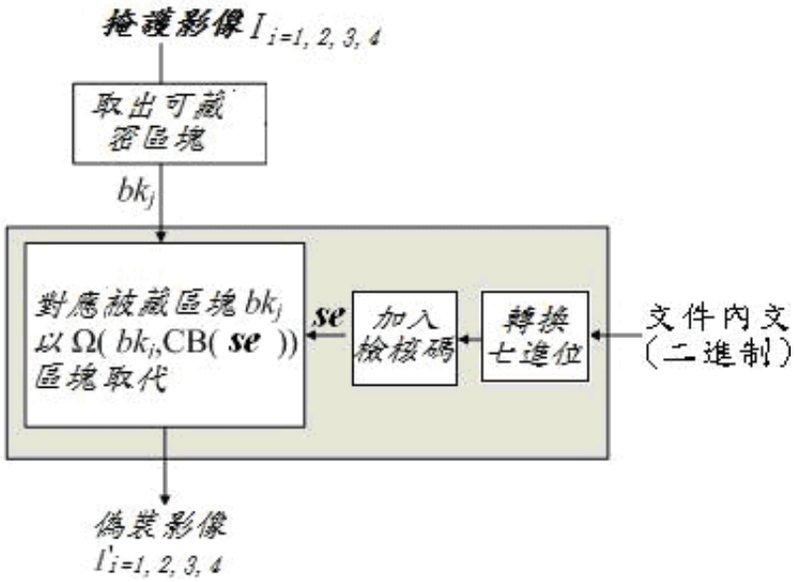


圖 6　藏密流程圖

操作步驟描述如下：

步驟 1。依序取出區塊 bk_j。

步驟 2。循序取出七進位值，以區塊Ω（bk_j，CB（*se*））取代。舉例如圖 7。

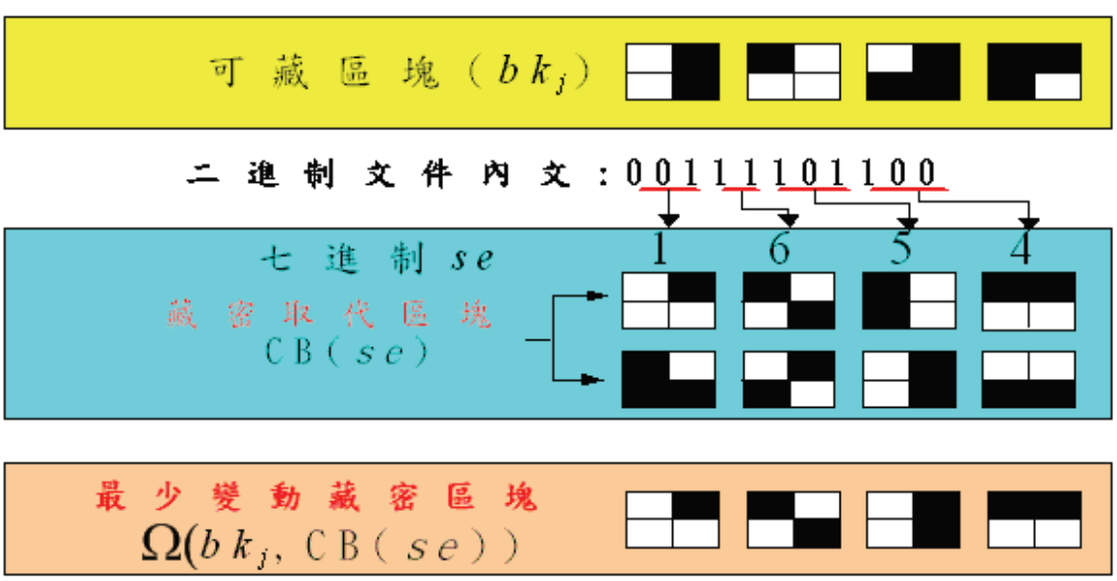


圖 7　藏密實例

二、竄改位置偵測與還原

本部份說明竄改位置偵測與文件內文還原兩部份，流程圖如圖 8。

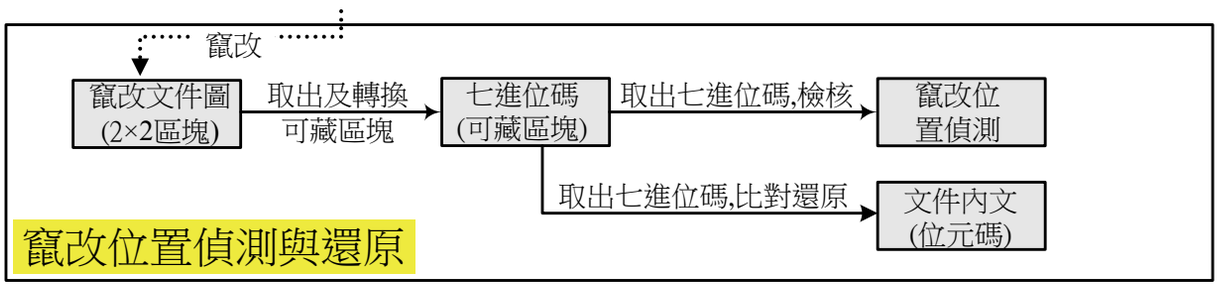


圖8　竄改位置偵測與還原流程圖

1、竄改位置偵測

竄改位置偵測步驟如下：

步驟 1：轉換可藏密區塊為七進位資料：將偽裝圖像均分成四等份 I_1', I_2', I_3', I_4'，並在四圖像中取出 2×2 可藏密區塊並轉換成七進位字串。

步驟 2：每四個七進位碼為一組，其中前面連續三個七進位數依（公式 1）計算若等於第四個七進位數字則通過檢核，反之若不相等則此區塊組必遭竄改將此區塊組標示出來。唯竄改後的圖形已與原圖不相同，必須從竄改部份後找出正確的原圖七進位順序才可順利偵測出取樣的正確性，如圖 9。

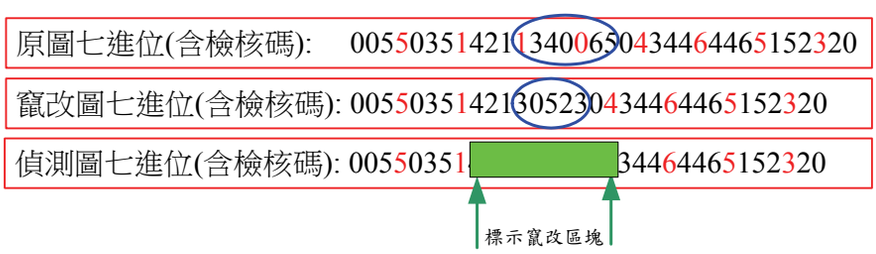


圖 9　文件竄改位置偵測與標示竄改區塊實例

由於竄改後取出的七進位長度也會改變，因此必須取出正確順序才可以保證未改變的部分正確無誤，因此從第一個錯誤檢核碼後需每次右移一位取出四個七進位依序檢核，直至正確為止。

2、文件內文還原

經過區塊組檢測完後，因為有四區塊藏四份原文七進位，因此以無誤部份為主，多數決的方法為輔取出正確的七進位，並去掉每組第四個檢測七進位，其餘字串則可取出並轉成二進位文件內文如圖 10。

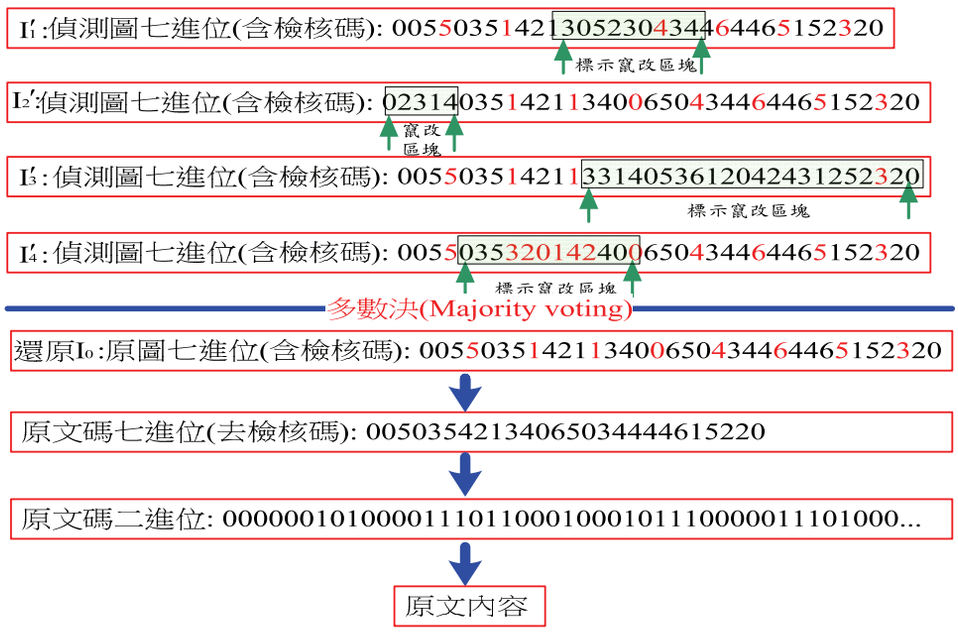


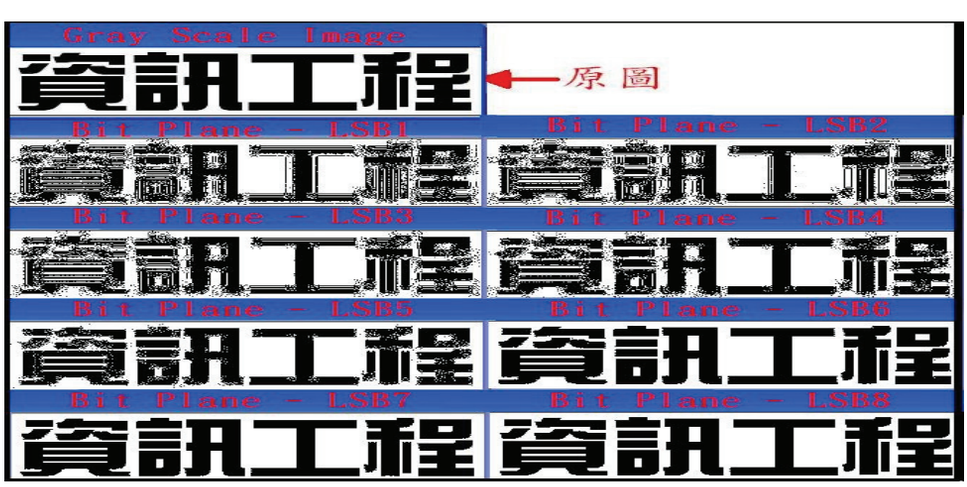
圖 10 取出原文內容實例

三、二元影像機制延伸應用於灰階、彩色圖像

可任意藏密於 LSB1~3（像素值 1，2，4）的灰階及彩色圖像[1, 4, 5, 7, 15]，其研究皆無法適用於僅能藏密於邊緣二元圖像，但二元影像藏密機制研究，可在不影響原灰階或彩色 LSB1~3 的藏密機制外，應用於 LSB4（像素值 8）bit plane 的灰階及彩色圖像，提供重要文件雙重的保護機制。

1、位元平面（bit plane）

灰階圖像中每一像素由 8 位元組成，分別取同位元則可組成 8 張二元圖像（Bit Plane）[1]，其中取出每個像素第 1 個位元所製成的二元圖像就是位元平面（LSB1），以此類推，取出第 2 個（LSB2）、第 3 個（LSB3）…總共可取出 8 張位元平面圖如圖 11。



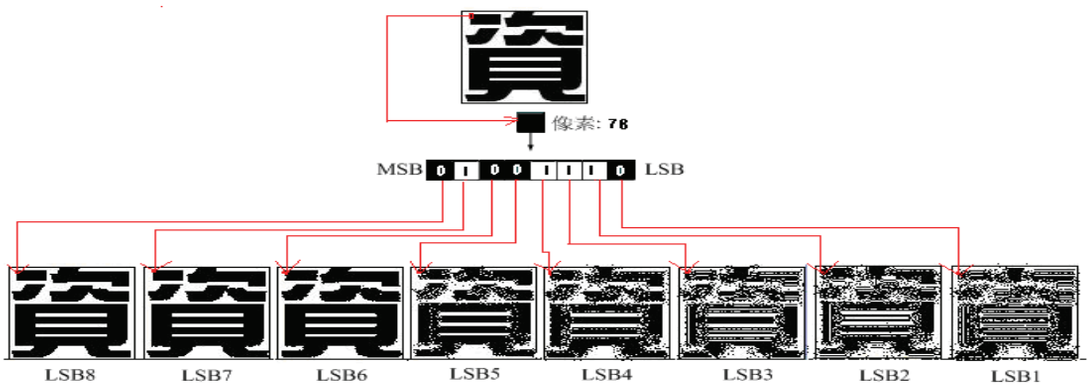


圖 11　取出 bit plane 的圖解

同樣地，彩色圖像由紅、藍、綠 3 組 8 位元像素組成，可分成 3 組 8 張二元圖像。

2.選取位元平面偽裝及藏密

許多竄改偵測位置與還原機制應用於灰階圖像中的 LSB，唯若破壞的部份同時發生於所有備份時，則無法達成復原目標，在不影響原來機制，視覺品質亦可兼顧的情況下，將文件內文藏數份於 LSB4 位元平面，除可由檢核碼偵測出竄改位置，亦可由正確及多數決的方式還原文件內文，提供雙重交叉的保護機制。彩色圖像因有 3 組 8 位元像素，採用交叉備份模式，更可大幅提高文件的保護。

肆、實驗結果

提出的機制主要著重於竄改偵測與復原因此評估的方式必須要滿足於相當的藏量視覺品質復原與竄改偵測，首先我們選擇一中文二元圖像尺寸 512×512 如圖 12（A）。

【借款契約書】 立借款契約書人王大明（以下簡稱甲方）、李小平（以下簡稱乙方），茲因借款事宜，訂立本契約書，條款如下： 一、 甲方願貸與乙方新台幣伍佰萬元正。 二、 借貸期限為一年，即自二零零九年八月三十日起至二零一零年七月十九日止。 三、 貸款年利率以二分計算，不得拖欠。 四、 遲延利息及逾期違約罰金加倍計算。 五、 本契約書之債權，甲方得自由讓渡予他人，乙方不得異議。 六、 乙方及保證人不依約履行時，願逕受法院執行，不得異議，因此所發生之費用悉由乙方及連帶保證人負擔。 七、 本契約一式五份，請求法院公證，除存案一份外，當事人各執乙份以備存查。 甲方:王大明　　乙方:李小平 西元二零零九年七月二十日	ADC9B4DAABB4ACF9AED1A5DFADC9B4DAABB4ACF9AED1A448A4FD A4A4A5FEA548A455 C2B2BAD9A5D2A4E8A7F5 AA59 AFC2A548A455C2B2BAD9A441A4E8AFF7A65DADC9B4DA A8C6A979AD71A5DFA5BBABB4ACF9AED1B1F8B4DAA670A455A440 A5D2A4E8C440B655BB50A441A4E8B773A578B9F4A4ADB855A4B8 BEE3A447ADC9B655B4C1ADADACB0A440A67EA759A6DBA445A451 A44BA67EA454A4EBA451A440A4E9B05FA6DCA445A451A445A67E A454A4EBA451A440A4E9A4EEA454A751AEA7A548B773A578B9F4 A843A6CAA4B8A4E9AEA7 A440A6CAA4B8AD70BAE2A441A4E8C0B3A9F3A843A4EBA451A4E9 B5B9A549A5D2A4E8A4A3B16FA9ECA4EDA57CBFF0A9B5A751AEA7 A4CEB94FB4C1B948ACF9BB40AAF7A8CCB773 A578B9F4A843A6CAA4B8A4E9AEA7A4ADA6CAA4B8AD70BAE2A4AD A5BBABB4ACF9AED1A4A7B6C5C576A5D2A4E8B16FA6DBA5D1C5FD B4E7BB50A54CA448A441A4E8A4A3B16FB2A7C4B3A4BBA441A4E8 A4CEAB4FC3D2A448A4A3A8CCACF9BC69A6E6AEC9C440A8FCAA6B B07CB0F5A6E6A4A3B16FB2A7C4B3A65DA6B9A9D2B56FA5CDA4A7 B64FA5CEB178A5D1A441 A4E8A4CEAB4FC3D2A448AD74BEE1A5D2A4E8A4FDA4A4A5FEA441 A4E8A7F5AA59AFC2B373B161AB4FC3D2A448B9F9 ADFB B1EAA4A4 B5D8 A5C1 B0EA A445 A451 A44B A67E A454 A4EB A451 A440 A4E9
（A）512×512 中文二元圖像原圖	（B）中文內碼

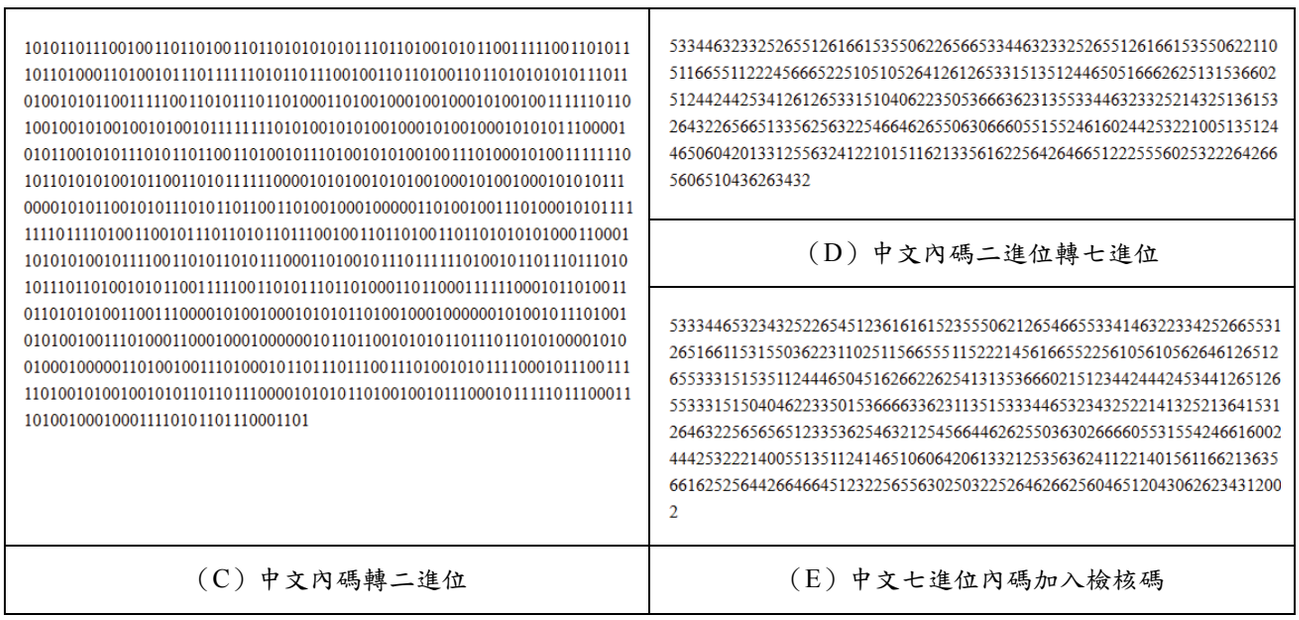


圖 12　512×512 中文二元圖像文件內文處理實作

由圖 12（A）中文轉成圖 12（B）中文內碼，並依序轉成二進位如圖 12（C），再轉成七進位如圖 12（D）。最後再每三個七進位加入一位七進位檢核碼如圖 12（E）。原圖均分成四等份，並從四等份中分別取出可藏區塊並轉成七進位數字串如圖 13（A），加入檢核碼的七進位

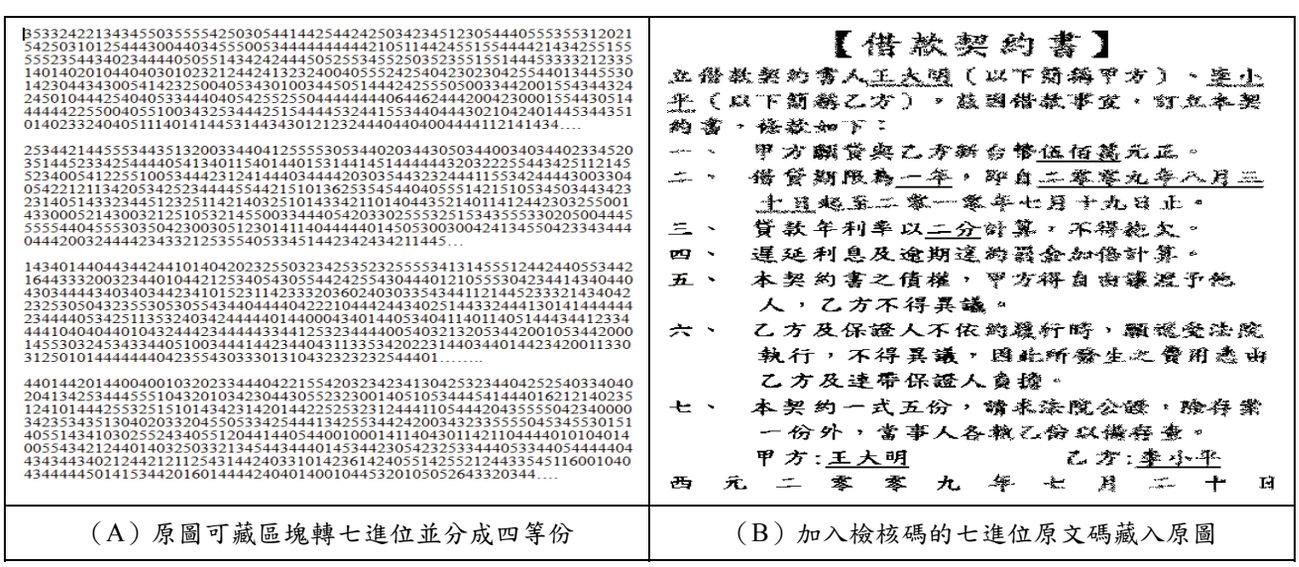
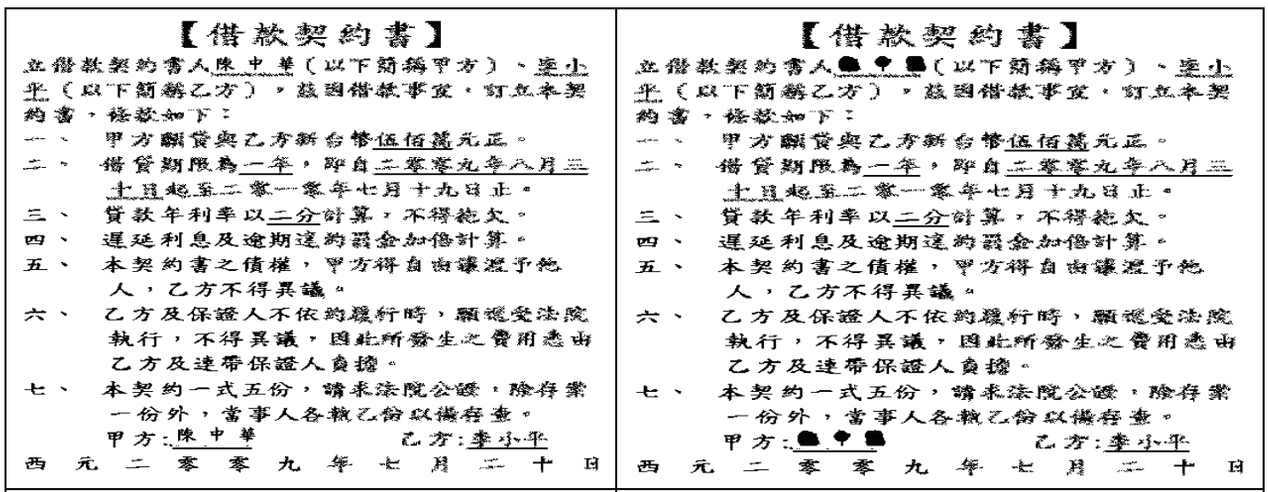


圖 13　加入檢核碼的七進位原文碼分別藏入原圖實作

原文碼分別藏入原圖取出的四份七進位數字串，並進一步將之藏回原圖中如圖 13（B）。若有人意圖竄改甲方王大明為陳中華如圖 14（A），則竄改位置部份藉由檢核碼及四份可藏密區塊比對可以確定竄改位置如圖 14（B）。



（A） 竄改甲方王大明為陳中華　　（B） 竄改位置偵測

圖 14　竄改位置偵測實作

我們除可偵測出竄改部分，並可借由所存入圖像文件的原文碼，萃取出原來的內容如圖 15。

借款契約書一立借款契約書人王大明以下簡稱甲方李小平以下簡稱乙方茲因借款事宜訂立本契約書條款如下一甲方願貸與乙方新台幣伍佰萬元正二借貸期限為一年即自二零零九年八月三十日起至二零一零年七月十九日止三貸款年利率以二分計算不得拖欠四遲延利息及逾期違約罰金加倍計算五本契約書之債權甲方得自由讓渡予他人乙方不得異議六乙方及保證人不依約履行時，願遲受法院執行，不得異議，因此所發生之費用悉由乙方及連帶保證人負擔七本契約一式五份，請求法院公證，除存案一份外，當事人各執乙份以備存查甲方王大明乙方李小平西元二零零九年七月二十日

圖 15　萃取原文內容

伍、結　論

本篇論文運用 2×2 的小區塊除可達成不易察覺的要求外，我們的設計更以互補區塊在二對一的最佳選擇下，在任何情況下每一區塊的變動量保證只有 0，1，或 2 位元的狀況發生，大幅的降低變動率，除此之外，在藏密率亦達到 2.857 bits/區塊（約 0.714）的高效率，由於區塊小，在藏密區塊相對增加的情況，藏密量大幅超前過去其它的研究，過去偵測還原機制只能侷限在灰階或彩色圖像上的研究。我們也藉由高藏量與低變動量來設計此偵測竄改位置與還原機制，在偵測還原機制上我們設計每三個區塊就加一個檢核區塊，並藏出多份的原文內碼因此除可很清楚的找出竄改的位置，多份的內碼也可協助我們刪除竄改部份或以多數決的方式找出正確七進位字串，進而萃取並還原原文之內文。

參考文獻

[1] Ambedkar, B., Mbedkar, B. “Steganography – Bit Plane Complexity Segmentation （BPCS） Technique,” International Journal of Engineering Science and Technology Vol. 2（9）, 2010.

[2] Chen, J., Chen, T. S. and Cheng, M. W., “A New Data Hiding Method in Binary Images, ” Proceedings of the Fifth International Symposium on Multimedia Software Engineering （ISMSE 2003）, pp. 88-93, Dec. 2003, Taichung, Taiwan.

[3] Fang, W. P., “Friendly progressive visual secret sharing, ” Pattern Recognition, Vol. 41, No. 4, pp.1410-1414, Apr. 2008.

[4] Fridrich, J, Goljan, M., and Du, R., “Detecting LSB steganography in color and gray-scale images,” IEEE Multimedia, vol. 8, no. 4, pp. 22–28, Oct.–Dec. 2001.

[5] Gao, Y. K., Li, X. L., and Lu, Y. F., “Detecting LSB Matching by Characterizing the Amplitude of Histogram, ” *Journal of Computers*, Vol 4, No 7, pp.646-653, Jul 2009

[6] Ho, Y. A. , Chan, Y. K. , Wu, H. C. , and Chu, Y. P., “High-capacity reversible data hiding in binary images using pattern substitution,” Computer Standards & Interfaces, Vol. 31, No. 4, pp. 787-794, June 2009.

[7] Ker, A. D., “Steganalysis of LSB matching in grayscale images,” IEEE Signal Process. Lett., vol. 12, no. 6, pp. 441–444, Jun. 2005.

[8] NBAFIPS PUB 46-1, “Data Encryption Standard,” National Bureau of Standards, U.S.Department of Commerce, Jan. 1988.

[9] Pamboukian, S. V. D., and Kim, H. Y., “Reversible Data Hiding and Reversible Authentication Watermarking for Binary Images,” Proceedings of the Sixth Brazilian Symposium on Information and Computer System Security （SBSEG’06）, Aug. 2006. Santos, Brazil. Available at http://www.lps.usp.br/~hae/sbseg2006-rdtc.pdf.

[10] Rivest, R., Shamir, A., and Adleman, L., “A Method for Obtaining Digital Signatures and Public-key Cryptosystems,” Communications of the ACM, Vol. 21, No. 2, pp. 120-126, Feb. 1978.

[11] Tsai, C. L., Chiang, H. F., Fan, K. C. and Chung, C. D., “Reversible Data Hiding and Lossless Reconstruction of Binary Images Using Pair-wise Logical Computation Mechanism,” Pattern Recognition, Vol. 38, No. 11, pp. 1993-2006, Nov. 2005.

[12] Tseng, Y. C., Chen, Y. Y. and Pan, H. K., “Data Hiding in 2-Color Images, ” IEEE Transactions on Computers, Vol. 51, No. 7, pp. 873-878, Jul. 2002.

[13] Wu, J. G., and Chung, K. L., “A New Binary Image Representation Logicodes,” Journal of Visual Communication and Image Representation, Vol. 8, No. 3, pp. 291–298, Sep. 1997.

[14] Wu, D.C., and Tsai, W. H., “Spatial-domain image hiding using an image differencing,” IEE Proc., Vis. Image Signal Process., vol.147,no.1, pp.29–37, 2000.

[15] Yang, H. and Kot, A. C., “Pattern-Based Data Hiding for Binary Image Authentication by Connectivity-Preserving, ” IEEE Transactions on Multimedia, Vol. 9, No. 3, pp. 528-538, Apr. 2007.

[16] Yang, S. H., and Liao, W. L.,“A compressed-domain watermarking scheme with the SPIHT coding,” Journal of Information Science and Engineering, vol. 26, no. 5, pp. 1755-1770, Sep. 2010.