改ざん検出機能を有する電子透かしの多層化方式

川島 康彰†　　満保 雅浩†　　岡本 栄司†
†筑波大学大学院システム情報工学研究科

1 はじめに

改竄検出には電子透かしが応用でき、幾つかの手法が提案されている [1][2]。 筆者らは保護領域で保護領域以外より高い確率で改竄を検出できる方式を文献 [3] で提案した。しかしこの方式では保護領域とそれ以外の領域という2つのクラスしか取り扱うことができなかった。そこで本論文では文献 [3] の方式を応用し、さらに多くのクラスに画像を分類し、クラスごとの画素の重要度に応じて改竄検出確率をより柔軟に割り当てられる方式を提案する。

2 基本方式

文献 [3] での方式は B^2 個の画素を1つのブロックとしてブロック毎に改竄検出を行う方式である。保護領域ブロックとそれ以外の領域のブロックを決定した後の各処理の概略は以下のようになっている。

保護領域のブロックに対する埋め込み処理

画素の各色成分の上位7ビットを連結し合計で $21B^2$ ビットを作る。このビット列をブロックごとに生成した暗号化鍵で暗号化し、透かし情報を生成する。この透かし情報を対象ブロック内の画素の各色成分のLSBに埋め込む。さらに保護領域以外のブロックからランダムに選らんだ領域を外部領域とする。そして、透かし情報から $uB^2(0 \le u \le 3)$ ビットをランダムに選別し、外部領域のLSBに埋め込む。

保護領域以外のブロックに対する埋め込み処理

全ての保護領域ブロックについて処理し終えた後、保護領域以外のブロックについて処理する。保護領域以外のブロックは各色成分の上位7ビットと保護領域のブロックから埋め込まれた情報を連結し $(21+u)B^2$ ビットを作る。このビット列をブロックごとの鍵で暗号化し、透かし情報を生成し、保護領域からの情報を埋め込んでいない色成分のLSBに埋め込む。

基本検出処理

対象ブロック内に含まれている透かし情報と再暗号化の出力を比較し改竄を検出する。一致する場合には改竄無し、不一致の場合には改竄有りと判断する。

保護領域における追加検出処理

保護領域に対してさらに改竄検出を行うため、外部領域にコピーしておいた透かし情報を回収する。もしコピーしておいた外部領域で既に改竄が判明していた場合には、その外部領域から回収せずに情報を破棄する。回収したビット列を多数決にかけ、最も多かったビット列と保護領域の透かし情報を比較し、基本検出処理と同様に一致、不一致で改竄検出を判断する。

3 基本方式の多層化

基本方式で説明したブロックを n 個 $(n \ge 1)$ のクラスに分類できる方式を以下に提案する。クラス番号は小さいほどクラスの重要度すなわち割り当てられる改竄検出確率が大きくなるようにする。

クラス1に対する埋め込み処理

クラス1に対して行う処理は基本方式での保護領域のブロックに対する処理と同じである。ただし外部領域として使用できるのはクラス2の領域のみとする。

クラス2～$n-1$に対する埋め込み処理

クラス2～nのブロックに対して行う処理は保護領域以外のブロックに対する処理と同様である。1つ小さいクラスからの情報を埋め込んでから、暗号化を行い対象ブロック内に埋め込む。基本方式と異なる点は2～$n-1$のクラスのブロックが1つ大きいクラスに情報を埋め込むことが出来るという点である。自分より1つ大きいクラスを外部領域として使用し、選択されたコピーを埋め込む。このコピーを埋め込む操作は保護領域のブロックに対する処理と同様である。

クラスnに対する埋め込み処理

最下位のクラス n は外部領域を使用せず、保護領域以外のブロックに対する処理と完全に同じになる。

基本検出処理

基本方式における基本検出処理と同様である。クラスnのブロックは透かし情報のコピーを持たないため、この基本検出処理しか行えない。

クラス1～$n-1$における追加検出処理

この処理は基本方式の保護領域における追加検出処理と同様の処理だが、クラス$n-1$から再帰的に検出を行う。下位クラスから再帰的に検出を行うことで、上位のクラスが回収する透かし情報のコピーの信頼性を上げることができる。

4 多層化方式の性能評価

各クラスにおける改竄検出確率

各クラスにおける改竄検出確率について考察する。最下位のクラス n における改竄検出では、対象ブロッ

Multiclass Digital Image Watermarking for Manipulation Detection
†Yasuaki KAWASHIMA　†Masahiro MAMBO　†Eiji OKAMOTO
†Graduate School of SIE, Univ. of Tsukuba

ク内に埋め込まれた透かし情報のみから改竄検出を行う。クラス n のブロック内に埋めこまれた対象ブロック自身の透かし情報は $(3-u)B^2$ ビットである。鍵情報を知らない改竄者がこれらのビット列を正しく生成できる確率は $(\frac{1}{2})^{(3-u)B^2}$ となる。

クラス $j(1<j<n)$ のブロックにおける検出は対象ブロック内に埋め込まれた透かし情報による検出とクラス $j+1$ の外部領域にコピーしておいた透かし情報での検出の2段階で行われる。対象ブロック内の情報による検出確率はクラス n と同等である。クラス j のブロックが外部領域にコピーしたビット列数を C_j、クラス j の改竄検出確率を Pd_j、外部領域として使用しているブロックに発生した改竄の個数を f とする。このとき回収した情報を多数決した結果が正しくなる十分条件は文献 [3] より $f<\frac{C_j}{Pd_{(j+1)}+1}$ である。$T_j=\frac{C_j}{Pd_{(j+1)}+1}$ とし、$f\geq T_j$ となる確率を $P(f\geq T_j)$ と表記する。このときクラス j の改竄検出確率 Pd_j は $1-(\frac{1}{2})^{(3-u)B^2}P(f\geq T_j)$ となる。この改竄検出確率はクラス n の改竄検出確率から再帰的に計算される。より多いコピー数を持つ上位クラスほど $P(f\geq T_j)$ が小さくなるため、より高い改竄検出確率を持たせることが出来る。

クラス1における改竄検出も同様にして計算できるが、クラス1のブロックは自ブロック内の埋め込み領域を全て自ブロックの情報で使用しているため $1-(\frac{1}{2})^{(3+u)B^2}P(f\geq T_1)$ となる。

誤検出に関する考察

ブロックが属するクラスに関する情報は秘密情報である。そのため意図的に狙ったクラスのブロックを改竄することは難しく、攻撃者にとっての有効な攻撃手段は誤検出の誘発となる。つまりランダムに改竄を行うことにより、多数決が正しくない結果を出し、誤検出が発生するように仕向ける攻撃である。この確率をできるだけ低くし、安全に使用するためには文献 [3] の考察より保護領域の面積は画像の $\frac{1}{3}$ 以下であることが望ましいことが分かっている。今回の方式でも、クラス j とクラス $j+1$ の面積にも同様の制約がある。このためクラス数を n と定めるとクラス1の領域として使用できる面積の最大値は画像サイズを MN としたとき、$\frac{MN}{B^2(2^n-1)}$ となる。すなわちクラス数 n とクラス1の面積には制限があり、クラス数もしくはクラス1の面積を決定すると、それによりもう一方のパラメータが制限される。ユーザは優先するパラメータを先に決定することになる。

5 実験結果

今回は $B=4$、$n=3$、クラス1からのコピー数を8、クラス2からのコピー数を7として実験を行った。改竄者が行う主な攻撃として、電子透かしの埋め込み直しと誤検出の誘発がある。電子透かしの埋め込み直しとは改竄を加えた後で、改竄が検出できないように電子透かしを埋め込み直す攻撃である。誤検出の誘発とは外部領域に改竄を加えることで検出対象ブロックに改竄が無いにも関わらず、改竄を検出させる攻撃である。今回の手法では、改竄者は各ブロックごとの鍵情報を知らないため、改竄後の透かし情報を計算することができず透かしの埋め込み直しを行うことが出来ない。そこで何らかの手法によって画素の変更を行い、誤検出を誘発させようとする。しかしどの画素がどのクラスに属するかというクラス情報は秘密として扱われるため、誤検出を発生させる効率的な手法を見出すことは困難であると考えられる。これにより改竄者が行う改竄はランダムノイズを付加することと同等と考えることが出来るようになる。そこで今回改竄としては10000個のランダムノイズを付加する改竄を選択した。この改竄検出を30回ほど繰り返し、平均を算出した。ブロックサイズ B を4に設定しているため、全クラスの改竄検出確率が高くなるので $B=1$ のときに比べクラス間の改竄検出確率の差が無いが、小さなクラスの方が高い検出確率を持っていることが分かる。また全体の改竄検出確率が高いため誤検出は全クラスにおいて0になっている。

クラス	改竄検出確率	誤検出
1	0.991824129	0
2	0.955492581	0
3	0.952933452	0

表1 改竄検出結果

6 まとめ

文献 [3] の方式を拡張し、クラスを n 個に分類し柔軟な改竄検出を実現できる方式を提案した。またその性能について理論的な考察と実験を行い、有用性を示した。今後の課題はセミフラジャイルな改竄検出機能の付加、公開鍵暗号系の利用、切り貼り攻撃に対する耐性の考察である。

参考文献

[1] 岩村恵市, 林淳一, 櫻井幸一, 今井秀樹, 安全な改ざん位置検出電子透かしに関する考察と提案, コンピュータセキュリティシンポジウム2001論文集, pp.283-288, (2001).

[2] M. M. Yeung and F. Mintzer, "An Invisible Watermarking Technique for Image Verification," IEEE Int. Conf. Image Processing, Vol. 2, pp. 680-683, (1997).

[3] 川島康彰, 満保雅浩, 岡本栄司, 柔軟な改ざん検出機能を有する電子透かし方式, 情報処理学会研究報告. CSEC, Vol.2007, No.48 pp. 19-24, (2007).