データハイディングによる改竄場所特定方法

利根川聡子[*1)]、上條浩一[*2)]、中村大賀[*2)]、森本典繁[*2)]、小出昭夫[*2)]
日本アイ・ビー・エム 株式会社 ソフトウェア開発研究所[*1)]、東京基礎研究所[*2)]

1. はじめに

データハイディング技術（電子すかし）は一般的に情報をデジタル画像に耐性を持って付加する技術だが、それに加え、変化に敏感な構造を持った埋込みを画像全体に施すことにより、フォーマット変換後にも画像認証をできるばかりでなく、画像に加えられた改変部分の特定をする事ができる。

一般的なデジタルデータの認証は一方向 Hash による電子署名が用いられる。しかしこの方法はファイルのフォーマット変換がなされただけで、電子署名がなされたデータであるかどうかの判定が不可能になるばかりでなく、フォーマット変換によるデータの改変と意図的な改竄を区別することもできない。

本稿は、画像全体に埋込まれた電子すかしの部分的擾乱を検知し、画像認証と意図的な改竄場所の特定に応用する方法について述べる。[1]

2. 電子すかしプロセス

電子すかしの埋込及び抽出プロセスを図2-1に示す。

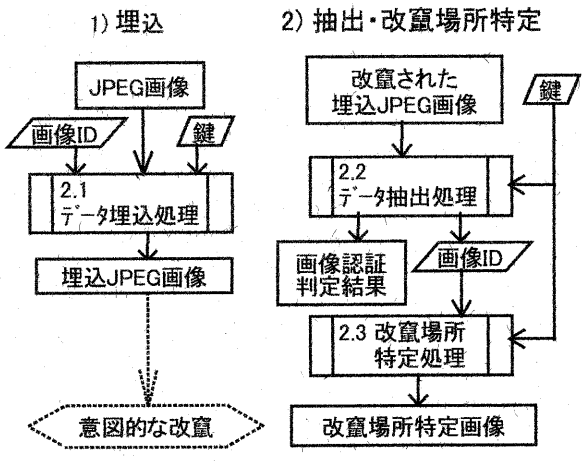

図 2-1　改竄場所特定プロセス

認証の対象になる画像は、デジタルカメラで生成された JPEG フォーマット画像が多いため、本稿では JPEG フォーマット画像へ直接データを埋込、抽出できる方法について説明する。

2.1 データ埋込処理

入力データとして JPEG フォーマット画像、鍵、埋込データ（画像 ID）がある。鍵から導かれる擬似乱数を用いて、画像 ID を繰り返し画像全体に配置する。JPEG をハフマン復号化し、DCT(Discrete Cosine Transform)係数を取り出した後、輝度成分変化させることによって埋込処理を行う。ハフマン符号化し JPEG フォーマットに復元して出力を行う。

詳細には、データ埋込は、隣接した２つの DCT ブロックからそれぞれ抽出された複数対の DCT 輝度係数ペアの大小関係を意図的に変化させることによって、'1'/'0'のビット値を表現させる。これは隣接した２つの同じ空間周波数の大小関係は、通常の量子化圧縮やフォーマット変換でもその性質を保持する特徴があるためである。また再圧縮による画像データの改変を防ぐため、意図的に変化させた複数対の DCT 輝度係数ペアの差分は十分に大きく取っている。

2.2 データ抽出処理

入力データとして第３者による意図的な改竄が行われた JPEG フォーマット画像、鍵がある。2.1 と同様に画像から DCT 輝度成分を取り出す。鍵から導かれる擬似乱数を用いて、画像 ID を DCT 輝度成分から抽出する。抽出した画像 ID と、入力画像が認証されたものであるか否かの判定結果（画像認証判定結果）を出力する。

詳細には、データ抽出は、隣接した２つの DCT ブロックからそれぞれ抽出された複数対の DCT 輝度係数ペアの大小関係から、2.1 章で行われた意図的な'1'/'0'のビット値表現を取り出し、埋込データと

---

[1] Alteration Detection using DataHiding
Satoko TONEGAWA,Koichi KAMIJO,Taiga NAKAMURA, Norishige MORIMOTO, Akio KOIDE
IBM Japan Ltd., Yamato Software Laboratory, Tokyo Research Laboratory

して抽出する。データは繰り返し埋込まれているので、部分的改竄やフォーマット変換による損失があっても抽出できる。

入力画像が認証されたものであるか否かの判定の詳細は、抽出されたビットの値と回数から判定する。データ埋込ビット数を a、埋込ビットの繰り返し回数を b とする。'1'/'0'のビット値表現状態と観測できた回数をm、それぞれの状態になる確率は 1/2 である。'1'（または'0'）のビット値表現状態がｘ回からm回観測される確率は、

$$Q = \sum_{i=x}^{m} mCi * (1/2)^m \quad \cdots(1) \quad 0 \leqq m \leqq b$$

となる。Q を 0.1 としたときのｘを求め、観測したビット値表現状態の回数ｎがｘ≦ｎであった場合、Q がａビット分すべて 0.1 となる確率は、

$$R = Q^a = 10E-a \quad \cdots(2)$$

ここでａが 12 ビットとすると、R=10E-12 となり、12 ビット全てが 0.1 以下の確率になることは非常にまれである。以上より、観測回数m中ビットの偏りがｎ個であるということがａビット全てに当てはまった場合、データ埋込処理がされている画像、つまり認証がされた画像であることが確認できる。意図的改竄が画像全体に及んだ場合、上記の判定が失敗するため「認証されていない画像」と判定される。

2.3 改竄場所特定処理

入力データとして、2.2 章で取り出された DCT 輝度成分と抽出した画像 ID がある。埋込処理によって意図的に作られた複数対の DCT 輝度係数ペアの大小関係は画像全体にわたって保たれている。この画像に対して部分的な改竄が行われた場合、この関係は改竄が行われた部分のみ破壊される。これを検知することによって、改竄場所を特定することができる。関係の破壊についての詳細を以下に述べる。通常、DCT 輝度係数ペア（a1, a2）の大小関係は、以下の(3)(4)が考えられる。 a1-a2=Δa

$\Delta a < 0$ または $\Delta a > 0$ …(3)

$\Delta a = 0$ …(4)

埋込処理直後の大小関係は画像全体のブロックが全て(3)状態になっている。一方通常の JPEG フォーマット画像は量子化処理の為、ほとんどのブロックが(4)状態になる。意図的に改竄が行われたブロックの大小関係も、ほとんどが(4)状態となる。つまり、以下 A),B) になっている DCT ブロックを、意図的に改竄されたブロックと判定できる。

A)推定される埋込ビット以外のビット値表現の状態

B)4)状態

判定したブロックにクラスタリング処理を施し、改竄場所として出力する。

このように、埋込処理で与えられた電子すかしは、フォーマット変換には耐性があるが、意図的な改竄には敏感であるという性質を持っている。

改竄場所特定処理により出力された改竄場所を、視覚的に表示した一例を以下に示す。左は改竄を施した画像、右画像の矩形で囲まれた部分が改竄場所を特定した部分である。

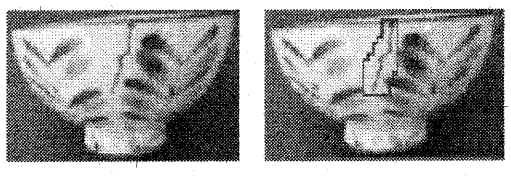

図 2-2 改竄場所特定例

3. まとめ

本手法を用いた自動車事故査定業務システム（文献[2]）のプロトタイプを作成し、実際の保険会社の査定用業務写真を用いたフィールドテストを行った。その結果、実用に耐える手法であることを確認された。今後、実用システムの構築を進める予定である。

参考文献

[1]上條、利根川　他“保険クレーム処理グループワークシステムにおけるデジタル写真の改ざん防止と検出機能(II)”

TECHNICAL REPORT OF IEICE, IN99-97, TM99-63, OFS99-50 (2000-01)

[2] 豊川、利根川　他“Secure Digital Photograph Handling Method with Watermarking Technique in Insurance Claim Process”　The International Society for Optical Engineering 2000/01/27