3M-07

フォーマット変換にロバストなデータハイディングの一手法

上條浩一
日本アイ・ビー・エム(株) 東京基礎研究所

1 はじめに

画像や音声等デジタルコンテンツに微小な変化を加え、ID 等の付加情報を埋め込み、検出する手法にデータハイディング (以下、DH) 技術があり、その応用の一つとして、デジタル写真の改竄検出がある。一般的方法は、対象となる画像の特徴量のハッシュ値 E を計算し、E を DH 技術を用いて画像自体に埋めこみ、検出時に、検出対象画像から導かれるハッシュ値と E とを比較し、改竄の有無を検出する、というものである。しかし、この改竄検出法では、微少な変化でハッシュ値が変わってしまうので、単なるフォーマット変換を施しただけで、人為的な改竄が無いのにも関わらず改竄有り、と判定するいわゆる false positive を引き起こしてしまう、という問題がある。この問題は、例えば JPEG→BMP 変換が行なわれた場合、BMP 画像において、R,G,B の画素値が上限、下限を超えたものを truncate する非線形のいわゆる overflow/underflow 処理が行なわれることによって引き起こされる。本論文では、JPEG 画像に埋めこみを行なう場合を例にとり、単なる JPEG→BMP 変換では、ハッシュ値は変化せず”改竄無し”と判定するが、微少な人為的改竄を施すと”改竄あり”と判定できる DH の一手法について報告する。

2 量子化ステップの性質を利用した改竄検出法

本方式では、埋めこみ対象 JPEG 画像をハッシュ値を埋めこむエリア $\boldsymbol{A}$ とそれ以外のエリア $\boldsymbol{B}$ に分け、$\boldsymbol{B}$ の画像の特徴量 (本論文では DCT の輝度成分の係数の量子化値) のハッシュ値 E を計算し、$\boldsymbol{A}$ に DH 手法を使って E を埋めこむ。このとき、$\boldsymbol{B}$ から計算されるハッシュ値が JPEG→BMP 変換で変化しない様に、JPEG の量子化ステップ (以下、Q 値) の性質を利用した”改竄マークを埋め込み”を $\boldsymbol{B}$ に対して行なう。

前提として、埋め込み者と検出者は、$\boldsymbol{A}$、埋め込み、検出に利用する複数の DCT 成分 d_l, ハッシュ関数 H とその鍵 K を共通に知っているものとする。

Datahiding method robust to format change
Koichi KAMIJO
IBM Japan Ltd., Tokyo Research Laboratory

2.1 埋め込み

図 1 に埋めこみ手順を示す。以下、この図に沿って、埋めこみ手順を説明する。

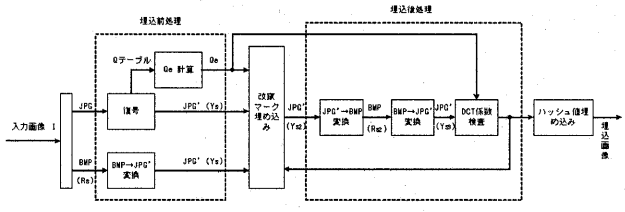

図 1: 埋めこみ手順

復号: 入力画像 I を復号化し、輝度、色差全ての 8×8 DCT 成分に対して、逆量子化を行う。以下、このように各 DCT 係数が逆量子化されている JPEG 画像フォーマットを JPG ’ フォーマットと呼ぶことにする。

Qel 計算: 埋め込みに使用する DCT 輝度成分 d_l の埋め込み Q 値 Qe_l を計算する。これは、埋め込み時に使用する Q 値であり、JPEG の量子化テーブルは書き変えない。いま、δ を、システム (decoder) の違いから生じる iDCT 計算時の最大計算誤差の 2 倍に設定し、I の d_l に対応する Q 値を Q_l とし、Qe_l を、

$$Qe_l = [(\delta - 1)/Q_l + 1]Q_l \qquad (1)$$

に設定する。

改竄マーク埋め込み: JPEG→BMP 変換が行なわれても量子化値が変化しないように、d_l の係数値を Qe_l の整数値倍になるような JPG' 画像を作る。具体的には、$c(i,k)$ を画像の i 番目の輝度 block(8×8 pixel) の DCT 成分 $k(0 \leq k < 64)$ の係数と定義し、

$$c(i,d_l) = nQe_l, \quad n = 0, \pm 1, \pm 2, ... \qquad (2)$$

が全ての i,l に対して成り立つように変化させる。

DCT 係数検査: JPG'→ BMP,BMP→ JPG' 変換を経て出来た JPG' フォーマット画像から得られる $c(i,d_l)$

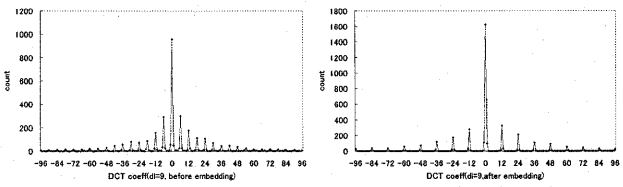

図 2: DCT 係数分布 (左:埋め込み前, 右:埋め込み後)

がある閾値 ϵ に対して

$$|nQe_l - c(i,d_l)| < \epsilon \text{ for } \exists n \qquad (3)$$
$$n = 0, \pm 1, \pm 2..$$

を全ての i,l に対して満たしているかどうかを検査する。ある block i が全ての l に対して式 (3) を満たしているとき、その block を、"安定化"している、と呼び、その block に対しての埋め込みは終了する。安定化していない block に対しては、"**改竄マーク埋め込み**" ルーチンに帰還し、埋め込みをやり直す。

帰還して改竄マーク埋めこみをやり直す場合、式 (2) の n として、$c(i,d_l)$ を Qe_l で量子化した値 ($\equiv n'$) を使うのでなく、n' より絶対値が小さい値を使うと $|c(i,d_l)|$ が 0 に近づき、再度 JPG'→BMP 変換を行なった後の overflow/underflow の影響が少なくなるが、n をあまり 0 に近づきすぎると画質が痛んでしまう。そこで、帰還を繰り返しながら徐々に $|n|$ を小さくしていく。帰還を繰り返し、全ての block が安定化したところで、I への改竄マーク埋め込み処理は終了する。

ハッシュ値埋めこみ: エリア $\boldsymbol{B}$ で DCT 成分 d_l の係数の Qe_l での量子化値のハッシュ値を計算し、$\boldsymbol{A}$ に埋める。埋めこみ方法として、"安定"な状態で、LSB 法を適応する方法がある。

2.2 検出, 検証

検出は以下の手順で行なわれる。以下は BMP 画像から検出する場合である。

埋込量子化ステップ逆算: 検査対象画像 I' から改竄検出を正しく行なう為には、検出側で Qe_l を知る必要があるが、I' が BMP 画像の場合、量子化テーブルを持たない。しかし、本方式では、埋め込み時に式 (1) を満たすように Qe_l を十分大きく取っている為、Qe_l を $\{d_l\}$ の DCT 係数のヒストグラムより逆算することが出来る。図 2 の右の図の例の場合、$Qe_l = 12$ と解る。

改竄検出 I' のエリア $\boldsymbol{B}$ の $\{d_l\}$ の係数から算出されたハッシュ値 E' と、$\boldsymbol{A}$ に埋めこまれているハッシュ値 E が一致するか否かで I' に人為的な改竄が行なわれたかどうかを判定する。

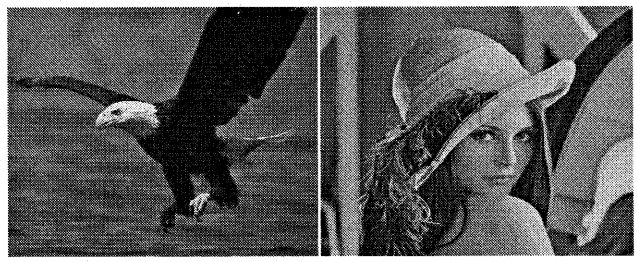

図 3: 実験画像例

検証 図 3 にある画像の他、デジカメで撮った、車、風景等様々な写真を使い、埋めこみを行ない、単なる JPEG→BMP 変換では"改竄無し"と判定するが、画像に点を加える等微少な改竄を行なうと、"検出あり"と判定することを確認した。

3 応用

本論文では、改竄の検出にフォーカスを当て、改竄場所の特定に関しては議論しなかったが、例えば、画像を"安定化"した状態で、DCT 係数の Qe_l での量子化値の LSB が規則性を持つよう変化させる事によって各 block に情報を埋め込みを行なえば、フォーマット変換にロバストだが、改竄検出と改竄場所の特定が出来る埋め込みも行なうことが出来る。

4 おわりに

本論文では、圧縮ドメインにおける Q 値の性質を利用した改竄検出方法を紹介し、その実験結果を報告した。本方式の大きな特徴は、改竄マークが overflow/underflow を伴うフォーマット変換に耐えうるが、微少の人為的改竄が検出が出来る事である。特に、人為的改竄が実際には BMP 等の非圧縮 domain で行なわれることを考えると、JPEG domain で改竄マークの埋め込みをされた画像が、BMP domain で改竄検出が可能であることは、とても有用な特徴である。

参考文献

[1] 上條浩一 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] 豊川和治 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" TECHNICAL REPORT OF IEICE,OFS99-29,IE99-38(1999-09)