2007－CSEC－37 (4)
2007／5／25

柔軟な改ざん検出機能を有する電子透かし方式

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

小切手の画像や交通事故写真など証拠能力を有する画像が痕跡を残さずに改竄できてしまうことは極めて問題である。このため改ざん検出に電子透かしを応用することが提案されているが、既存の方式では画素ごとの改竄検出確率が全画素で均等である、もしくは、どの画素においても1または0のみであり、画素ごとの重要度が改ざん検出確率に反映されていなかった。そこで本研究では各画素における改竄検出確率の指定を画素の重要度に応じてより柔軟に行える方式を提案し、保護領域に対する改ざん検出確率が保護領域以外より高くなることを理論解析と実験を通して示す。

Digital Image Watermarking Detecting Manipulations with a High Probability

Yasuaki Kawashima　Masahiro Mambo　Eiji Okamoto
Graduate School of SIE, Univ. of Tsukuba

Since a digital evidental image requires high accuracy, any manipulation to the image should be detected certainly. Digital watermarking is a promissing method for detecting manipulations. However, known manipulation detection watermarking methods provide only limited type of distribution of detection probability, e.g. equal detection probability in all pixels, or probability only 1 or 0 in any pixel. In contrast to the known methods, we propose a manipulation detection watermarking method which has a probability distribution reflecting the degree of importance of each pixel. Theoretical analysis and conducted experiments show our proposed method achieves high probability of manipulation detection for an important pixel area.

1　はじめに

近年、インターネットが急激に普及したことにより、デジタルコンテンツを取り扱う様々な新しいビジネスが立ち上がってきた。デジタルコンテンツはコピーしても劣化することが無いため、インターネットを介した配布や販売に適している。しかし編集も容易なため、改竄されやすいという欠点も持ち合わせており、コンテンツの不正配布により著作権が侵害される恐れもある。これらの欠点はコンテンツビジネスの発展を妨げており、改竄や不正配布抑止技術の整備は急務である。

特に、交通事故現場の写真や領収書の画像など証拠能力を有する画像が、改竄されてしまうことは非常に問題である。2003年3月31日にはロサンゼルスタイムズに改竄された写真が掲載されるという事件も実際に起きている。この改竄は記者が自白することにより発覚したが、自白が無ければ、発覚することは無かったと言えるほど巧妙なものであり、単にデジタルコンテンツ作成者のモラル低下を防ぐ手立てを講ずるということだけでは済まない。何らかの技術的対策が必要であることを示唆している。つまり改竄があった場合には、改竄検出の必要性の度合いに応じて、的確に改竄を検出できるようにするべきである。この改竄検出に電子透かしが応用できることが知られている。

電子透かしを用いた改竄検出では、脆弱な電子透かしを埋め込んでおき、検出できないときに改竄があったとみなす。この脆弱な電子透かしとして、画像のパリティを制御する方式とパリティを制御する代わりに認証子を埋め込む方式の2つがある。改竄を隠蔽される危険性という点から見ると、後者のほうが安全である。しかし後者の方式では筆者の知る限り、指定された領域だけで高い改竄検出確率を提供する方式 [2] と全画素で均等な改竄検出確率を提供する方式 [1] [3] しか存在しない。画像には保護すべき領域とそれ以外の領域が存在するため、全画素で均等な改竄検出確率では、保護しなくてもよい領域に必要以上に高い改竄検出確率を割り当てているという点で効率的ではない。また保護すべき領域だけを保護する方式では保護の必要性の度合いは低いが全く保護しなくてよい訳ではない領域の改竄を検出することが出来ない。そこで本論文では、認証子を用いた安全性の高い電子透かしにより改竄検出を行う際に、ユーザによって各画素の竄検出確率を画素の重要度に応じて指定できる方式を提案する。

2 関連研究

認証子を用いた電子透かしにより画像の改竄を検出する試みとして暗号化を用いる Mintzer らの方式 [3] やその改良方式である岩村らの方式 [1] と Hash 関数を用いる田中らの方式 [2] などが提案されている。

田中らの方式

この方式では、まず保護する領域を設定する。次に保護する領域をブロックに分割する。各ブロックの画素の上位 7 ビットを全てのブロックについて連結し Hash 関数の入力とする。対応する出力結果の先頭 160 ビットにブロックの位置情報 16 ビットを加え埋め込み情報とする。画像に対してラプラシアンフィルタを用いてエッジ検出を行い、エッジ部分に埋め込み情報を埋め込んでいく。検出操作時には Hash 関数を用いて埋め込み情報を再生成し、抽出された埋め込み情報と照らし合わせて改竄を検出する。

この手法では、保護する領域に関してはほぼ 1 に近い改竄検出確率を持つが、保護領域以外では改竄検出確率が 0 となってしまい、柔軟な改竄検出が実現できていない。また保護領域以外が改竄されたことが原因でビットの不一致が発生したときも保護領域の改竄と誤って判定してしまう欠点がある。加えて、ハッシュ関数は公開されるため、誰でも埋め込み情報を計算できてしまい、埋め込み直しを行おうとする攻撃者により有利となる情報を与えていることになる。

岩村らの方式

この方式では、まず一つの乱数シードから画素固有の暗号化鍵を生成する。各画素の青以外の色成分と青の上位 7 ビットを連結し合計 23 ビットのビット列を作成し、そのビット列を画素固有の鍵で暗号化する。暗号化結果の下位 1 ビットを青の最下位ビットに置き換える。検出操作時には、再び画素固有の鍵を生成し、青の最下位ビットと比較する。不一致の場合には改竄あり、一致する場合には改竄無しと判定する。

改竄を加えようとする攻撃者は画素固有の鍵を知らないため、田中らの方式と異なり、埋め込み情報である青の最下位ビットを知ることが出来ない。そのため攻撃者が改竄した結果、青の最下位ビットが正しい結果になっている確率は $\frac{1}{2}$ となる。しかも、全画素に関して 1 ビットの埋め込み情報を埋め込んでいるので、改竄検出確率は全画素において均等に $\frac{1}{2}$ となる。しかし画像における重要度は画素により異なり、保護したい画素と保護しなくてもよい画素が存在する。このことから保護しなくてもよい画素の改竄検出率を保護したい領域に振り分けられる方がが好ましいと言える。

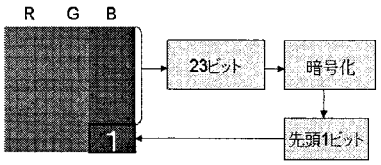

図 1 岩村らの方式

3 提案方式

本論文では原画像の状態を変更するあらゆる処理を改竄として取り扱う。また改竄検出を行う画像はビットマップ形式や PPM 形式のような非圧縮画像とする。鍵情報として、暗号化鍵生成用乱数、埋め込み位置決定用乱数、ブロックサイズ、保護ブロックの位置情報が必要になる。暗号化アルゴリズムは任意である。

以下に保護する領域とそれ以外の領域を指定して、保護領域に高い改竄検出確率を割り当てる方式を示す。

埋め込み処理

まず埋め込み処理の流れについて説明する。図 2 に保護領域の埋め込みに関するモデルを示す。

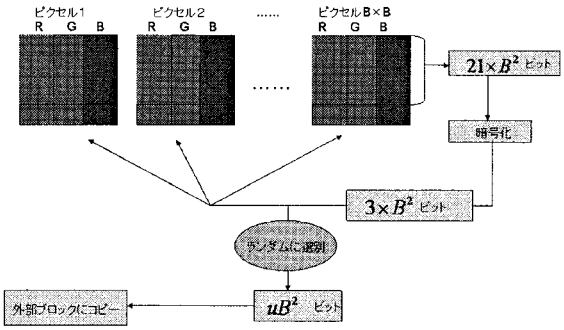

図 2 保護領域の埋め込み処理

STEP 1 任意の対象画像をピクセル数 B × B のブロックに分割する。ただし B は自然数。

STEP 2 保護領域とするブロックを指定する。

STEP 3 保護領域に指定された各ブロックの RGB 成分の上位 7 ビットを連結した $21 \times B^2$ ビットを暗号化し、埋め込み情報をブロック毎に作成する。

STEP 4 埋め込み情報として暗号文の先頭 $3B^2$ ビッ

トを取り出し、保護領域自身の最下位ビットに埋め込む。

STEP 5 保護領域の埋め込みデータから $\lfloor uB^2 \rfloor$ ビットをランダムに選別する。但し、u は $0 < u < 3$ を満たす実数。

STEP 6 埋め込み位置を決定するための乱数を選ぶ。この乱数を用いて、保護領域に指定されていないブロックの中から埋め込み用のブロック（以後、単に外部ブロックと呼ぶ）を決定し、STEP5 で選別した uB^2 ビットのデータを埋め込む。この外部ブロックの決定とデータ埋め込み処理を複数回繰り返す。

STEP 7 保護領域以外の全てのブロックにおいて、STEP6 で保護領域の情報を埋め込まなかった色成分の上位 7 ビットと埋め込んだ色成分の 8 ビットを連結した $(24-u)B^2$ ビットの暗号化を行う。

STEP 8 STEP7 で暗号化した情報の先頭 $(3-u)B^2$ ビットを取り出し、保護領域以外のブロックの保護領域の埋め込み情報を埋め込まなかった最下位ビットに埋め込む。

検出処理

次に改竄検出処理の流れについて説明する。まず図 3 に改竄検出処理の流れを示す。また図 4 に誤検出防止処理に関するモデル図を示す。

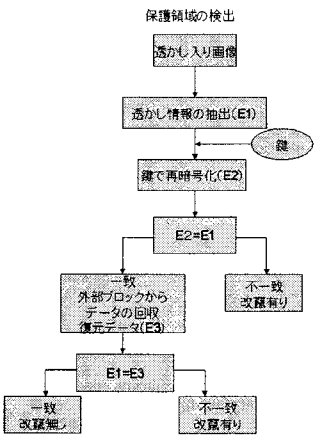

図 3 保護領域の検出処理手順

STEP 1 保護領域のブロック、保護領域以外のブロックそれぞれに対して、埋め込み処理のときと同様に暗号化し、暗号文を最下位ビットと比較することで改竄があるか否か判定する。一致する場合は改竄無し、不一致の場合には改竄有りと判定する。

STEP 2 保護領域のブロックのうち改竄が無かったと判定されたブロックに関しては、対応する外部ブロックに埋め込んでおいた情報を回収し、更にチェックする。

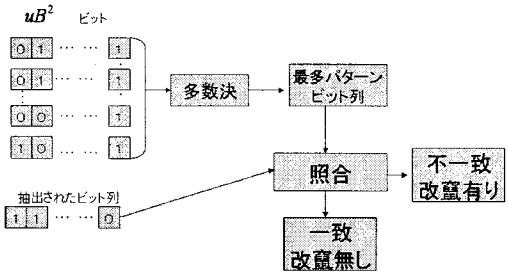

図 4 誤検出防止処理

STEP 3 改竄が無いと判断された外部ブロックごとに uB^2 ビットの情報が回収され、その中で最も多いパターンのビット列を改竄の無いビット列と判断する。

STEP 4 決定されたビット列と保護領域ブロックの最下位ビットを比較し、一致しなかった場合には改竄有りと判断する。

保護領域の埋め込み情報を外部ブロックに埋め込むとき、どの外部ブロックのどの色成分を用いるかは秘密の乱数により決定する。乱数を用いず、外部ブロックと保護領域ブロックの対応付けを攻撃者が知ることが出来るようにしてしまうと外部ブロックを改竄することにより、保護領域自体には改竄が無いにも関わらず、改竄があるかのように誤検出させることが可能となってしう。このため、4.1 節に示すようにこの対応付けが成功する確率が十分に小さくなるようにパラメータ設定を行う必要がある。

保護領域と保護領域以外に分割することで改竄検出確率の分布の柔軟化を実現しているが、この領域分割を複数回繰り返し、さらに細かい改竄検出確率を指定することが可能となる。

4 提案方式の性能評価

4.1 検出確率に関する考察

本提案方式では、画像に加えられる加工は全て検出すべき改竄と定義している。改竄検出に対する特有の攻撃として透かしの埋め込み直しと外部ブロックの改竄による誤検出の誘発がある。

透かしの埋め込み直しは、改竄とつじつまが合う透かしを埋め込み直して改竄を隠蔽する攻撃である。本方式では各ブロックの暗号化に用いる暗号化鍵を秘密

の乱数にすることで攻撃者は正しい埋め込み情報を求めることが出来ない。これにより埋め込み直しへの攻撃耐性が向上しているといえる。具体的な確率は本節で算出する。

外部ブロックの改竄による誤検出の誘発とは、外部ブロックを改竄することで、改竄の無かった保護領域のブロックに改竄があったかのように見せかける攻撃である。本方式では、外部ブロック自体でも改竄検出を行うことや保護領域と外部ブロックの対応関係を秘密にすることで、この確率を低く抑えている。具体的な確率は本節で算出する。

画像サイズは $M \times N$ ピクセル、保護領域として指定されるブロックの個数を p 個、ブロックサイズを B^2 ピクセルとする。このとき、画像中のブロック数は $\frac{MN}{B^2}$ 個、保護領域の 1 ブロックあたりの使用できる外部ブロック数は $\frac{MN-B^2p}{B^2p}$ 個と表現することが出来る。保護領域以外のブロックに関しては合計 $(3-u)B^2$ ビットのビット列をそのブロック自身に埋め込んでいる。また保護領域では、保護領域のブロック自身に埋め込んだビット列は $3B^2$ ビットである。よってそのブロック自身に埋め込んだビット列により改竄を検出できる確率は以下のようになる。

$$\begin{cases} 1-(\frac{1}{2})^{(3-u)B^2} & (\text{保護領域以外})\cdots(1) \\ 1-(\frac{1}{2})^{3B^2} & (\text{保護領域})\cdots(2) \end{cases}$$

保護領域で $(\frac{1}{2})^{3B^2}$ の確率で改竄の見逃しがあった場合には外部ブロックに埋め込んでおいた埋め込み情報のコピーからビット列を復元し、埋め込み情報と照らし合わせ、改竄を検出する。このときの改竄検出率を導出する。コピーを埋め込める外部ブロックは最大 $\lfloor\frac{MN-B^2p}{B^2p}\rfloor$ 個である。そのうちの f 個のブロックが改竄されていると仮定する。コピーが埋め込まれた外部ブロックで改竄が無ければ、必ず改竄が無いと判断され回収される。改竄があった f 個のコピーうち、平均で $f(\frac{1}{2})^{(3-u)B^2}$ 個は改竄検出をすり抜け、改竄が無かったと判断され回収されてきてしまう。これにより改竄が無かったと判断され回収されてくるコピーの数は $(\lfloor\frac{MN-B^2p}{B^2p}\rfloor - f) + f(\frac{1}{2})^{(3-u)B^2}$ のように表現できる。ビット列の復元は多数決によって決められているので、$(\frac{MN-B^2p}{B^2p} - f) > f(\frac{1}{2})^{(3-u)B^2}$ すなわち $f < \frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}$ となるときは正しいビット列が復元される。このときは改竄を検出できる確率が 1 となり、保護領域ブロックで改竄が無いにも関わらず改竄を検出する誤検出率が 0 となる。$f > \frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}$ のときには、正しくないビット列が復元されてしまう。コピーの埋め込み位置はブロック、およびブロック中の埋め込みビット位置とも乱数によって決定されているため、復元されるビット列はランダムなビット列となってしまい、改竄者が意図するビット列を復元させることは出来ず、改竄を検出できる確率は $1-(\frac{1}{2})^{uB^2}$ となる。しかしランダムなビット列が復元されてしまうことにより、改竄が無いときでも改竄と判定してしまう誤検出が発生し、その確率は $1-(\frac{1}{2})^{uB^2}$ となる。

保護領域のブロックで改竄の見逃しが有ったとき、ビット列の復元による改竄検出確率をまとめると以下のようになる。

$$\begin{cases} 1 & (f \le \frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}\text{のとき}) \\ 1-(\frac{1}{2})^{uB^2} & (f > \frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}\text{のとき}) \end{cases}$$

$\frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}$ を T とおき、$f \le T$ の発生する確率を $P(f \le T)$、$f > T$ の発生する確率を $P(f > T)$ とする。保護領域ブロックの改竄検出確率 P_d をまとめると次のように表現することが出来る。

$$\begin{aligned} P_d &= \left(1-(\tfrac{1}{2})^{3B^2}\right) \\ &\quad + (\tfrac{1}{2})^{3B^2}\Big(P(f \le T) + (1-(\tfrac{1}{2})^{uB^2})P(f > T)\Big) \\ &= 1-(\tfrac{1}{2})^{(3+u)B^2}P(f > T) \end{aligned}$$

また保護領域のブロックで改竄が無かったときに誤って改竄を検出してしまう確率をまとめると以下のようになる。

$$\begin{cases} 0 & (f \le \frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}\text{のとき}) \\ 1-(\frac{1}{2})^{uB^2} & (f > \frac{MN-B^2p}{B^2p\{(\frac{1}{2})^{(3-u)B^2}+1\}}\text{のとき}) \end{cases}$$

さらに 3 章で述べたように保護領域と外部ブロックの対応付けが成功しないようにする必要がある。保護領域 1 ブロックに対応する外部ブロックの半数が見つけられてしまうと、外部ブロックの改竄により誤検出を引き起こされてしまう。ここで、保護領域以外に出来るだけ多くのコピーを埋め込む場合を考える。どのブロックが保護領域に指定されているかという情報を改竄者が知っていると仮定したときに、半数の外部ブロックを見つけることができる確率は次のように表すことができる。

$$\frac{(\frac{MN-pB^2}{pB^2})C(\frac{MN-pB^2}{2pB^2})}{(\frac{MN-pB^2}{B^2})C(\frac{MN-pB^2}{2pB^2})} = \frac{3}{\sqrt{2}} - \frac{MN}{\sqrt{2}pB^2}$$

この確率を閾値 t 以下であると仮定したとき、次の式が導かれる。

$$pB^2 \le \frac{MN}{3-\sqrt{2}t}$$

またこの式において t を 0 に近づけると次のようになる。

$$\lim_{t\to 0}\frac{MN}{3-\sqrt{2}t} = \frac{MN}{3}$$

pB^2 は保護領域に指定されている画素の個数である。すなわち保護領域に指定される画素数が $\frac{MN}{3}$ のときには保護領域と外部ブロックの対応付けが成功する確率は 0 になることが分かる。

4.2 既存方式との比較

本方式のメリットとして、ブロック化することで保護領域とそれ以外の領域の両方で改竄検出率を高められることが挙げられる。また、埋め込み情報生成用鍵を用いたことにより田中らの方式と比べて埋め込み直しに対する攻撃耐性が向上し、埋め込み位置情報生成用鍵を用いたことにより岩村らの方式と比べて埋め込み直しに対する攻撃耐性が向上したこともメリットである。

一方、外部ブロックの使用により岩村らの方式では 0 であった誤検出発生確率が田中の方式と同様に増加するというデメリットがある。このデメリットに関しては、多数決を行うことにより、田中らの方式と比べ誤検出確率は低く抑えることが出来ると予想される。またブロック化することにより改竄位置の特定能力が低下するというデメリットもあるが、改竄検出能力は向上する。

4.1 節の式 (1)(2) より、保護領域とそれ以外の領域の改竄検出確率の際が最も開くのは $B = 1$ のときである。提案方式では $B = 1$ のとき、ブロック単体による改竄検出確率は保護領域で $\frac{7}{8}$、保護領域以外で $\frac{1}{2}$ である。岩村らの方式では $\frac{1}{2}$ だった改竄検出確率と比べると保護領域での改竄検出率が向上していることが分かる。

4.3 実験

実験条件

今回実装したプログラムでは $B = 2$ とし 2×2 のブロックに分割し、電子透かしを埋め込んだ。採用した暗号はバーナム暗号である。実験には 128x128 画素の画像を用いた。現実的な攻撃としては透かしの埋め込み直しによる改竄隠蔽と外部ブロックの改竄による誤検出の誘発があるが、本提案方式は保護領域に関する情報と暗号化鍵、保護領域ブロックと外部ブロックの対応関係が秘密情報であるため、外部ブロックの改竄によって誤検出を発生させようとしても、どの画素を改竄すればいいのかという情報が得られない。また改竄を隠蔽するために透かしの埋め込み直しを行おうとしても、暗号鍵は秘密情報なので入力部に対応する正しい埋め込み情報が分からない。これらいずれの攻撃もランダムノイズを加える攻撃と同等の効果があると言うことができる。そのため本実験ではランダムノイズを付加する攻撃を採用した。対象とする画像 lena を図 5 に示す。保護領域として始点座標 (40,40)、終点座標 (79,79) の領域を指定した。保護領域を示した画像を図 6 に示す。黒くなっている部分が保護領域である。また電子透かしの入った画像を図 7 に示す。

図 5 原画像

図 6 保護領域

図 7 透かし入り画像

図 8 ノイズ付加画像

実験結果

図 7 に対して 1000 個のランダムノイズを付加した画像が図 8 である。図 8 に対して改竄検出を行った画像が図 9 である。白くなっている点は改竄が検出されていることを表す。また原画像を 1000 個の点で RGB を 1 階調づつ変化させるという人間が検知できないレベルで変更した場合の例を図 10 に示す。ランダムノイズを負荷した場合と同様に改竄を検出できていることがわかる。1000 個のランダムノイズを付加

図 9 改竄検出結果

図 10 改竄検出結果

する実験を乱数シードを変更しながら100回行い、改竄を検出できた点の個数を算出し平均を求めた結果、表1のような結果が得られた。

表1　実験結果 (lena)

領域	誤検出（個）	改竄検出確率
保護領域	27.2	0.928
保護領域以外	0	0.914

誤検出は誤検出を起こしたブロックの個数である。

表1で分かるように保護領域での改竄検出確率は、保護領域以外の改竄検出確率よりも高くなっていることが分かる。また保護領域と保護領域以外の改竄検出確率の差が比較的小さいようにも思われるが、4.1節で示したようにそれぞれの改竄検出確率がブロック化サイズ $B \times B$ に依存しているため B の値を小さくすることにより、より保護領域と保護領域以外の検出確率の差は大きくなると予想でき、$B = 1$ のとき、差が最大になる。

5　おわりに

本論文では保護する領域とそれ以外の領域を指定して透かし埋め込む手法を提案し、理論解析と実装を通して、高い検出確率を実現していることを確認した。また画像に改竄を加え、実際に改竄が検出できることの検証を行った。

今後の課題として改竄の度合いが大きいときのみ改竄検出を行う方式の検討、誤検出率の抑制、保護領域指定の自由度向上が挙げられる。今回の手法では拡大縮小も改竄と位置づけているため改竄として検出されるが、画像の意味は把握できる改竄であるため、改竄は無かったと判断する方が良い場合もある。また左右上下のいずれかに1画素分だけ平行移動させられると全体の改竄として検出されるという問題点もある。本方式でも誤検出を抑制する工夫は施したが、それでも誤検出は発生している。この誤検出を抑えることは重要な課題と言える。また改竄検出確率の指定をさらに自由に行える方式の考案も課題である。埋め込み容量の制限を考慮した適切なの埋め込みビット数割り当ても検討する必要がある。具体的には、一枚の画像に対して埋め込めるビット数には $3MN$ ビットという制限があり、各ブロックに対する埋め込みビット数の割り当てを工夫する必要がある。

今回提案した手法では、コピーしたビット列の多数決によって埋め込みビット列の復元を行ったが、これは概念的には誤り訂正符号の手法であると捉えられる。電子透かしを用いた改竄検出に適した誤り訂正符号の設計も今後の課題である。

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] 田中愛子, 岡本栄司, 小野束,　電子透かしを用いたデジタル画像改ざん検出法に関する研究, 筑波大学大学院博士課程システム情報工学研究科修士論文, (2004).

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

[4] 小野束,「電子透かしとコンテンツ保護」, オーム社 (2001).

[5] 松井甲子雄　,「電子透かしの基礎　-マルチメディアのニュープロテクト技術-」, 森北出版株式会社 (1998).

[6] 小松尚久, 田中賢一,「電子透かし技術　デジタルコンテンツのセキュリティ」, 電機大出版局 (2004).