

電子透かしにおける最適な検出について

吉田真紀

大阪大学大学院情報科学研究科

Maki Yoshida

Graduate School of Information Science and Technology, Osaka University,

1-5 Yamadaoka, Suita-shi, Osaka 565-0871, Japan

Email: maki-yos@ist.osaka-u.ac.jp

1 はじめに

電子透かしは、デジタルコンテンツの著作権を保護するための技術である。透かしとして、コンテンツの所有者や購入者の情報などが埋め込まれる。例えば、透かしとして購入者情報を埋め込んでおくことで、不正コピーが押収されたとき、透かしを検出（あるいは抽出）することでコピーの流出元が特定でき、その結果不正コピーを抑止できる。検出結果（抽出結果）の誤りは流出元の誤認や見逃しにつながるため、その信頼性について数学的な根拠を与えることは重要であり、これまでに様々な研究が行われている。本稿では、画像に対する検出型の電子透かしの信頼性保証に関する研究 [1, 2, 4, 5, 9, 14, 15, 16, 17] を紹介する。以降では、透かしと埋め込み領域の値をそれぞれ透かし信号とホスト信号と呼ぶ。また、透かし入りコンテンツと検出対象コンテンツをそれぞれ透かし入り信号と検出対象信号と呼ぶ。

透かし信号の検出は、与えられた検出対象信号に特定の透かし信号が埋め込まれているか否かを判定する処理である。透かし信号の検出において、二種類の誤りが起こりうる。検出対象信号に透かし信号が埋め込まれていないにも関わらず、その透かし信号を検出する誤り（誤検出）と、埋め込まれているにも関わらず、その透かし信号を検出しない誤り（見逃し）である。誤検出の方が、より重大な問題を引き起こす場合が多い。例えば、透かし信号としてコンテンツ購入者情報を埋め込んだ場合、誤検出は流出元の誤認につながり、正当な購入者に疑いがかかる。よって、誤検出確率が十分小さいことが求められるが、両方の確率はトレードオフの関係にあるため、誤検出確率を小さくし過ぎると見逃し確率が大きくなってしまう。

よって、[1, 2, 4, 5, 9, 14, 15, 16, 17] では、誤検出確率を指定された確率とした上で見逃し確率を最小とすること（最適な検出）を目標としている。最適な検出法の設計は、埋め込み領域と埋め込み処理を特定した上で行われる。これまでに対象となっている埋め込み領域は画像の画素値や周波数成分である。また、埋め込み処理は、比較的簡単な埋め込み規則である加法規則か乗法規則に従う埋め込み処理である。

我々の最近の研究成果 [8, 10, 11, 12] として、一般的な相関型電子透かしに共通して適用可能な誤検出確率保証法がある。最適な検出 [1, 2, 4, 5, 9, 14, 15, 16, 17] との違いは、埋め込み領域と埋め込み規則を限定しないこと、既存の電子透かしに容易に適用可能とするために、検出処理に変更を加える必要がないことである。提案法によって、誤検出確率は指定した確率とできるが、検出処理に変更を加えないため、見逃

し確率が最小となるとは限らない。

以下では、まず2章で最適な検出に関する既知の結果を示し、3章で我々の誤検出確率保証に関する結果を示す。

2 最適な検出に関する研究

2.1 最適な検出法の分類

最適な検出法は透かし信号の埋め込み領域と埋め込み規則で分類される。埋め込み領域は、画素空間と周波数領域に分けられる。画素空間に埋め込む場合は画素の輝度値やRGBを変更し、周波数領域に埋め込む場合は画像を周波数成分に変換した上で周波数成分を変更する。周波数領域のホスト信号として離散フーリエ変換 (DFT), 離散コサイン変換 (DCT), 離散ウェーブレット変換 (DWT) 係数が考えられている。それぞれの長所と短所を示す。

- 画素空間
 - 長所: 埋め込み, 検出処理が高速
 - 短所: 攻撃に対する耐性が低い
- 周波数領域
 - 長所: 攻撃に対する耐性が高い
 - 短所: 埋め込み, 検出処理が低速

近年、計算機の性能が向上したため、攻撃に対する耐性が高い周波数領域への埋め込みが主流となっている。

埋め込み規則には、加法規則と乗法規則の二つがある。加法規則は、埋め込み領域の値に透かしを足しこむ規則であり、乗法規則は、埋め込み領域の値と透かしを掛け合わせた結果をもとの値に足しこむ規則である。それぞれの規則を直観的に記述すると以下のようになる。

- 加法規則: [透かし入り信号] = [ホスト信号] + [透かし信号]
- 乗法規則: [透かし入り信号] = [ホスト信号] × (1 + [透かし信号])

乗法規則の方は、埋め込みによる変更量が埋め込み領域の値に依存するため、見た目の変化 (品質の劣化) が少ないという長所をもつ。

表 1: 既存の最適な検出法の分類

	画素空間		周波数領域	
	輝度値	RGB	DFT	DCT, DWT
加法規則	VS98[17]	SVCS99 [15]	–	CH01 [4], SWL03 [16]
乗法規則	–	–	BBRP01 [1], BBRP03 [2]	HKC03 [9], CH03 [5]

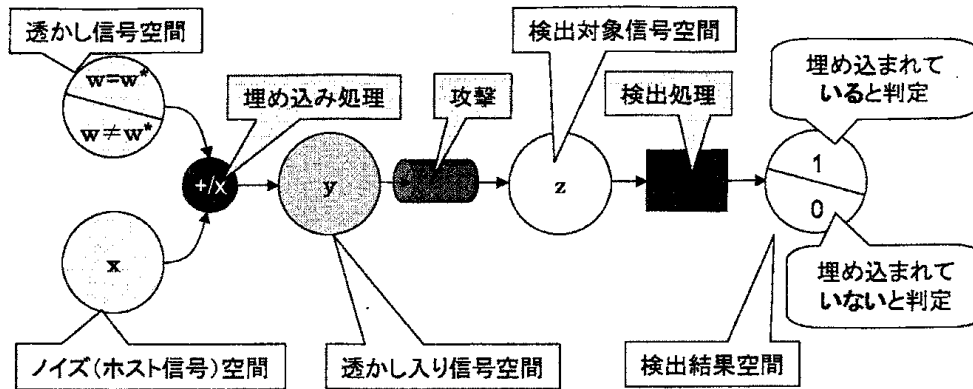


図 1: 透かし信号 w^* の検出システム

表 1 に、これまでに提案された最適な検出法をまとめる。ある埋め込み領域と埋め込み規則に対しては最適な検出法が二つ提案されている。その違いを述べる。

- DFT と乗法規則に対する最適な検出法

BBRP01 の検出法は一つの透かしが埋め込まれる場合を、BBRP03 は複数の透かしが埋め込まれる場合を対象としている。

- DCT, DWT に対する最適な検出法

CH01 (加法規則を対象) と HKC03 (乗法規則を対象) の検出法は攻撃がない場合を対象としており、攻撃がある場合は最適となることを保証しない。一方、SWL03 (加法規則を対象) と CH03 (乗法規則を対象) の検出法は攻撃がある場合でも (攻撃の強さについてある条件が満たされれば) 最適となることを保証する。

2.2 最適な検出法に関する結果

透かし信号を検出する際、ホスト信号は透かし信号に対するノイズとみなされる。図 1 に透かし信号 w^* の検出システムを示す。検出システムでは、埋め込まれる透かし信号 w として、検出対象の透かし信号 w^* か異なる透かし信号が選ばれ、埋め込み処理によってホスト信号 x がノイズとして付加される。その結果得られる信号が透かし入り信号 y であり、 y に攻撃が付加された結果得られる信号が検出対象信号 z である。検出処理では、透かし信号 w^* と検出対象信号 z から計算される検出値 $v(w^*, z)$ を閾値 λ と比較し、以下のように行われる。

- $v(w^*, z) > \lambda$ ならば、 z に w^* が埋め込まれている (1) と判定する。
- $v(w^*, z) \leq \lambda$ ならば、 z に w^* が埋め込まれていない (0) と判定する。

最適な検出法に関して、以下の三点が研究されている。

- ホスト信号（ノイズ）の振る舞いを表す確率分布の選択
ホスト信号の振る舞いは、どのような確率分布を用いてモデル化すればよいか
- 最適な検出を実現する検出値と閾値の公式を導出
 - 攻撃がない場合を対象とするならば、透かし入り信号の分布に対して、どのような検出値と閾値を用いれば最適な検出が可能か、
 - 攻撃がある場合を対象とする場合は、攻撃についてある前提をおいた上で検出対象信号の分布を決め、どのような検出値と閾値を用いれば最適な検出が可能か、

なお、ホスト信号の分布と埋め込み規則が決まれば透かし入り信号の分布は決まる。

- 検出値と閾値の計算アルゴリズムの設計

最適な検出を実現する検出値と閾値を透かし信号と検出対象信号からどのように計算すればよいか

本節の残りでは、ホスト信号の振る舞いを表す確率分布と、最適な検出を実現する検出値と閾値の公式について既知の結果を紹介する。

ホスト信号の振る舞いを表す確率分布として、表現能力が高く、数学的に扱いやすい分布が利用される。

- 画素空間

輝度値と RGB は Gaussian 分布

- 周波数領域

DFT 係数は Weibull 分布. DCT, DWT 係数は Generalized Gaussian 分布

最適な検出を実現する検出値と閾値の公式は、透かし信号の検出を仮説検定とみなすことで、Neyman-Pearson の補題から導出できる。Neyman-Pearson の補題は、良い検定方式を構成する上で重要な補題であり、有意水準（第一種の過誤）に対して第二種の過誤を最小とする棄却領域（棄却点）を与える。よって、透かし信号の検出を、“ z に w^* が埋め込まれていない” という仮説の検定とみなせば、第一種の過誤が起きる確率（有意水準）が誤検出確率 P_{fp} に、第二種の過誤が起きる確率が見逃し確率に対応する。これにより、誤検出確率 P_{fp} 、透かし信号 w^* 、検出対象信号 z に対する最適な検出を実現する検出値 $v(w^*, z)$ と閾値 λ は以下の公式で与えられる。

$$v(w^*, z) = \frac{f_z(z|w = w^*)}{f_z(z|w \neq w^*)}$$

$$P_{fp} = P(v(z) > \lambda | w \neq w^*)$$

ここで、 $f_z(z|\cdot)$ は検出対象信号 z の条件付き確率分布、 $\overline{P_{fp}}$ は指定された誤検出確率である。

公式の中に検出対象信号の条件付き確率分布が現れるため、このままでは検出値も閾値も計算できない。よって、ホスト信号の確率分布に対して妥当な前提をおき、比較的容易に計算可能な検出値の公式を導出することが課題となる。例えば、埋め込み領域が画素空間で埋め込み規則が加法規則の場合 [15, 17]、最適な検出を実現する検出値は以下の式で与えられる。

$$Cor(w^*, z) = \frac{1}{n} \sum_{i=1}^n w_i z_i$$

ここで、 $z = (z_1, z_2, \dots, z_n)$, $w^* = (w_1, w_2, \dots, w_n)$ である。 $Cor(w^*, z)$ は、検出対象信号やホスト信号の確率分布を推定する必要がないため、計算が容易で、利便性が高い。ただし、埋め込み領域や埋め込み規則が異なれば最適な検出を保証しない。

最適な検出を実現する検出値を用いた場合と用いない場合では見逃し確率に大きな差がでる。例えば、誤検出確率を5%に指定し、埋め込み領域を周波数領域 (DCT, DWT 係数)、埋め込み規則を加法規則とした場合、最適でない検出法 ($Cor(w^*, z)$ を検出値として用いた方式) の見逃し確率20%に対して、最適な検出法 CH01 [4] の見逃し確率はわずか0.5%となる。

3 我々の最近の研究

我々の研究目的は、既存の一般的な画像用相関型電子透かしに共通して適用可能な誤検出確率保証法を設計することである。そのため、電子透かしであらかじめ定義されている検出値を変更することはせず、閾値を選択することで誤検出確率を保証する。まず、3.1節で対象とする相関型電子透かしのモデルについて説明した後、誤検出確率保証法として提案した閾値選択法 [10, 11] について3.2節で簡単に紹介する。提案した閾値選択法では検出対象コンテンツの原コンテンツが必要となるが、比較的単純な動画用電子透かしへ適用する場合には原画像を不要とすることができる [12]。3.3節では不要とするための方法を紹介し、3.4節で提案した閾値選択法の有効性に関する評価結果を示す。

3.1 相関型電子透かし

相関型電子透かしシステムは、2.2節の検出システムに従うが、用いられる埋め込み規則や検出値は特に限定しない。埋め込み規則を $embed$ と表し、ホスト信号 x と透かし信号 w に対して $embed$ を適用した結果得られる透かし入り信号を $embed(x, w)$ と表す (すなわち、 $y = embed(x, w)$)。ただし、検出値に対して、以下の式 (1) および式 (2) が十分高い確率で満たされるものとする。

$$v(w, x) < v(w, embed(x, w)) \quad (1)$$

$$v(w, embed(x, w')) < v(w, embed(x, w)) \quad (2)$$

二つの式は、埋め込まれた透かし信号との検出値の方が埋め込まれていない透かし信号との検出値より大きいことを表す。

$embed$ と v の例として静止画像用のパッチワーク法 [3] における埋め込み規則 $embed_p$ および検出値 v_p を挙げる。パッチワーク法は、埋め込み領域として画素空間を対象としている。パラメータとして透かしの強度 δ および変更する画素数 $2N$ をもつ。

- 埋め込み規則 $embed_p$ に従った埋め込み処理
 1. w を用いて擬似乱系列を生成する。
 2. この擬似乱系列を用いて x から N 個の画素の組 (a_i, b_i) ($1 \leq i \leq N$) を選択する。
 3. 全ての i において a_i の輝度値を δ だけ増加させ、 b_i の輝度値を δ だけ減少させる。
- 検出値 $v_p(w^*, z)$ の計算

1. $embed_p$ と同じ方法により, w^* を用いて z から N 個の画素の組 (a_i, b_i) ($1 \leq i \leq N$) を選択する.
2. (a_i, b_i) から以下の式によって統計量 $v_p(w^*, z)$ を求める.

$$v_p(w^*, z) = \frac{1}{N} \sum_{i=1}^N (a_i - b_i)$$

なお, 検出対象信号 z が攻撃を受けていない場合, z に w^* が埋め込まれていないならば, $v(w^*, z)$ は 0 を中心に分布し, z に w^* が埋め込まれているならば ($z = embed_p(x, w^*)$ ならば), 2δ を中心に分布する.

3.2 閾値選択法

文献 [10, 11] では, 誤検出確率を保証可能とする閾値を選択可能とするために, 多くの実用的な相関型電子透かしにおいて共通して満たされうる性質を導出し利用する. 導出した性質は次の性質 (P) である.

(P) 検出対象信号 z に検出対象透かし信号 w^* が埋め込まれていないならば, 式 (3) を満たす透かし信号ペア (w_1, w_2) が存在する. ただし $w^* \neq w_1$ であり $w^* \neq w_2$ である.

$$\frac{v(x, w_1) - v(x, w^*)}{v(x, w_2) - v(x, w^*)} = \frac{v(z, w_1) - v(z, w^*)}{v(z, w_2) - v(z, w^*)} \quad (3)$$

この性質は, 多くの相関型電子透かしが基本とする方式であるパッチワーク方式をもとに導出した. コンテンツとして静止画像, 攻撃として JPEG 圧縮・伸長を用いた場合に, この性質が満たされることを, [11] で実験的に示している.

性質 (P) を閾値選択法でどのように利用したかを述べる. まず, 透かし信号の検出を “性質 (P) が満たされる” という仮説の検定とみなす. すなわち, 仮説を棄却することが z に w^* が埋め込まれていると判定することとする. そして, 有意水準を指定された誤検出確率 $\overline{P_{fp}}$ とした検定における棄却点を検出の閾値 λ として出力するようにする. この閾値を用いて検出を行うことで, 透かし信号が埋め込まれていないとき (性質 (P) が満たされていないとき) に, 埋め込まれているとする (仮説を棄却する) 確率が指定された確率となる.

なお, 性質 (P) にホスト信号 x が現れるため, 閾値選択法は入力として検出対象信号 z と透かし信号 w^* だけでなく, ホスト信号 x も必要となる. 一般に閾値選択法の利便性を考えるならば, ホスト信号を必要としない方が望ましい. そこで [12] では, 動画像における隣接フレームの類似性を利用することで, 単純な動画像用相関型電子透かしにおいて閾値選択法を利用する際, ホスト信号を不要とすることを考えた. 次節では, その不要とする方法を簡単に示す.

3.3 動画像用電子透かしへの適用法

文献 [12] において, 閾値選択法の適用先とした単純な動画像用相関型電子透かし法を示し, ホスト信号 (動画像) を不要とする方法を示す. 動画像は, 複数のフレームが時系列に従って並ぶことで構成されている. 適用先とした動画像用相関型電子透かし法は 3.1 節で示したモデルに従い, 動画像中の一部のフレームに対して, 静止画像用の相関型電子透かしをそのまま適用する単純なものである. ただし, 連続したフレ



図 2: Entrance Hall における一フレーム

ムには埋め込まないとする。動画像の検出値は、埋め込み対象となったフレームから計算される検出値によって決まるとする。

閾値選択法を適用するためには、検出対象フレームのホストフレームが必要となる。動画像には、隣接するフレームは類似しているという性質がある。文献 [12] では、このフレーム間の類似性を利用して、検出対象フレームの隣接フレームをホストフレームの代わりとして用いることでホストフレームを不要としている。

3.4 評価結果

提案した閾値選択法の有効性の評価結果を紹介する。評価では実際の誤検出確率が指定した確率に近いかなかを調べている。相関型電子透かしとして静止画像用電子透かしと動画像用電子透かしの二つが適用対象としてあるが、ほぼ同じ結果が得られているため、ここでは動画像用電子透かしに適用した場合の結果を示す。その際、様々なフレーム間類似度に対する結果を得るため、検出対象フレームの隣接フレームだけでなく、様々な強さの攻撃を加えることで類似度を下げた隣接フレームをホストフレームの代わりとして用いる。

まず評価の条件を示す。

- 原動画像

映像情報メディア学会¹の標準動画像の一つである Entrance Hall (図 2 参照) を用いた。この動画像は、720×486 ピクセルのフレーム 450 枚で構成される。

- 相関型電子透かし: フレームへの透かしの埋め込みおよび検出に、パッチワーク法を用いた単純な動画像用電子透かしを用いた。パッチワーク法のパラメータは、[3]において原フレームの品質を保ち、かつその上で透かしの検出誤りを低くするのに十分であるとされているものを用いた。
- 鍵: 32 ビットの整数。
- 動画像への攻撃: ビットレートが 5Mbps の MPEG 圧縮・伸長を用いた。MPEG 圧縮・伸長には、mpeg2encode と mpeg2decode (MPEG Software Simulation Group²)を用いた。

¹<http://www.ite.or.jp/>

²<http://www.mpeg.org/MPEG/MSSG/>

- フレーム間類似度を下げたための攻撃: ノイズ付加, StirMark4.0 [6, 7] による回転, StirMark3.1 をそれぞれ単独で用いた.
- 指定した誤検出確率 $\overline{P_{fp}}$: 文献 [12] での評価と同じ 0.01.
- 類似度の尺度: 2 枚のフレームの類似度を測るために, 以下の式 (4) で計算される相互相関を用いた.

$$\frac{\sum_{k=1}^K (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^K (x_k - \bar{x})^2 \sum_{k=1}^K (y_k - \bar{y})^2}} \quad (4)$$

ここで, x_k および y_k は 2 枚のフレームそれぞれにおける各画素の輝度値を表し, \bar{x}, \bar{y} はそれぞれ x_k, y_k の平均を表す. K はフレームあたりの画素数である. 相互相関は, $[-1, 1]$ の値をとり, 2 枚のフレームの類似度が高いほど 1 に近い値となり, 類似度が低いほど 0 に近い値となる. また濃淡は反転しているが類似度が高い場合は, -1 に近い値となる.

3.5 評価結果

評価結果を表 2 に示す. 表 2 は, ホストフレームの代わりに利用した各フレームについて, 検出対象フレームとの類似度 (相互相関), 誤検出確率の平均を示している. 誤検出確率は, 類似度が低くとも指定した確率 0.01 から大きく離れていない. すなわち, 誤検出確率は保証されているといえる. ただし, 類似度が低くなると見逃し確率が上がり, ホストフレームの代わりに利用した各フレームについて, 検出対象フレームとの類似度 (相互相関) が 0.9 以下では, 見逃し確率と誤検出確率の和がほぼ 1 に近い値になっている. これは, 指定した誤検出確率でランダムに “埋め込まれている” と検出する場合と違いがないことを意味する. これは, 評価で用いた相関型電子透かしの埋め込み領域が, 攻撃に対する耐性が弱い画素空間となっているためである. 攻撃に対して強い相関型電子透かしを評価に用いることで, 見逃し確率は小さくなると考えられる.

表 2: Entrance Hall における相互相関と誤検出確率

相互相関	誤検出確率
0.98	0.0105
0.92	0.0101
0.89	0.0102
0.84	0.0108
0.79	0.0116
0.73	0.0091
0.71	0.0107
0.55	0.0114
0.37	0.0113

4 おわりに

本稿では、画像に対する検出型の電子透かしの信頼性保証に関する研究として、最適な検出に関する既知の結果 [1, 2, 4, 5, 9, 14, 15, 16, 17] と我々の最近の研究成果 [8, 10, 11, 12] を紹介した。電子透かしの信頼性を数学的に保証することは重要な課題である。攻撃がない場合を対象とした最適な検出法は多く提案されており、多くの結果が得られているといえる。しかし、攻撃がある場合を対象とした最適な検出法はまだ少ない。また、近年は最適な検出に関する結果をもとに、最適な抽出を実現することが考えられている。よって、電子透かしの信頼性保証に関する今後の課題として、攻撃がある場合も、攻撃の強さについてなんらかの条件が満たされるならば、その信頼性について数学的に保証可能とする検出法や抽出法の設計が考えられる。

参考文献

- [1] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "A New Decoder for the Optimum Recovery of Nonadditive Watermarks," *IEEE Trans. Image Processing*, vol. 10, pp. 755-766, May. 2001.
- [2] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Optimum Decoding and Detection of Multiplicative Watermarks," *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp. 1118-1123, April. 2003.
- [3] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM System Journal* vol.35, no.3&4, pp. 313-336, 1996.
- [4] Q. Chen and T. S. Huang, "An Additive Approach to Transform-Domain Information Hiding and Optimum Detection Structure," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 273-284, Sept. 2001.
- [5] Q. Chen and T. S. Huang, "Robust Optimum Detection of Transform Domain Multiplicative Watermarks," *IEEE Trans. Signal Processing*, vol. 51, no. 4, pp 906-924, April. 2003.
- [6] F. A. P. Petitcolas, R. J. Anderson, and Markus G. Kuhn, "Attacks on copyright marking systems," *Information Hiding, LNCS1525*, pp.219-239, 1998.
- [7] F. A. P. Petitcolas, "Watermarking schemes evaluation," *Proc. IEEE Signal Processing*, vol.17, no.5, pp.58-64, 2000.
- [8] 藤田 高彬, 岡本 邦宏, 吉田 真紀, 藤原 融, "動画像用相関型電子透かしの誤検出確率を保証可能とするフレーム間類似度の評価," *SCIS2005 予稿集*, pp.1033-1038, 2005.
- [9] Y. Hu, S. Kwong, and Y. K. Chan, "The Design and Application of DWT-Domain Optimum Decoders," *Digital Watermarking, LNCS2613*, pp.22-30, 2003.
- [10] 岡本 邦宏, 上野 貴之, 吉田 真紀, 藤原 融, "統計量利用型電子透かしに対する検出結果の信頼性保証," *SCIS2004 予稿集*, pp.879-884, 2004.
- [11] K. Okamoto, T. Ueno, M. Yoshida, and T. Fujiwara, "A Method to Ensure Reliability of a Detection Result for Correlation Based Watermark Detection Scheme," *Proc. ISITA2004*, pp.299-304, 2004.

- [12] K. Okamoto, T. Fujita, M. Yoshida, and T. Fujiwara, "Correlation Based Watermark Detection Ensuring Given False Positive Error Probability for Video Using Inter-Frame Similarity," CSS2004 論文集, pp.163-168, 2004.
- [13] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on copyright marking systems," Information Hiding, LNCS1525, pp. 219-239, April. 1998.
- [14] F. A. P. Petitcolas, "Watermarking schemes evaluation," Proc. IEEE Signal Processing, vol. 17, no. 5, pp. 58-64, September. 2000.
- [15] E. Sayrol, J. Vidal, S. Cabanillas, and S. Santamaria, "Optimum Watermark Detection in Color Images," ICIP99, val.2, pp. 231-235, Oct. 1999.
- [16] Y. Shao, G. Wu, and X. Lin, "Optimal Detection of Transform Domain Additive Watermark by Using Low Density Diversity," Digital Watermarking, LNCS2613, pp. 105-112, 2003.
- [17] J. Vidal and E. Sayrol, "Optimum Watermark Detection and Embedding in Digital Images," Proc. IEEE Multimedia Signal Processing, pp. 285-290, Dec. 1998.
- [18] 山本 基夫, 汐崎 陽, 岩田 基, "フレーム間の類似性を利用した動画像用相関型電子透かし," SCIS2004 予稿集, pp.1215-1220, 2004.