可視型電子透かしに関する提案とその安全性の評価

片岡　謙一朗　　岩村　惠市

東京理科大学
102-0073　東京都千代田区九段北 1-14-6

kataoka@sec.kagu.tus.ac.jp

あらまし　可視型電子透かしには，可逆性（秘密情報により可視透かしを除去可能），透明性（原画像が確認可能），安全性（秘密情報無しに透かしを除去・修正不能）という要求がある．それに対し，従来の可視型電子透かしは可逆性や透明性を実現できたとしても，安全性に関してはそのアルゴリズムが秘密であることを根拠としている場合が多く，十分な考察が行われている手法は少ない．本論文では，共通鍵を用いて上記可視型電子透かしの全要求を満たす手法を提案する．また，その結果を実験によって確認する．

Proposal and Evaluation of the Security of a Visible Watermarking

Kenichiro Kataoka　　Keiichi Iwamura

Tokyo University of science
1-14-6 Kudankita,Chiyoda-Ku Tokyo,102-0073,Japan

kataoka@sec.kagu.tus.ac.jp

Abstract　In general, there are three major requirements for visible watermarking. Visibility(the watermark must be easily identified), Transparency(the watermark must not significantly obscure the image details beneath it), Security(the original pixels in the watermarked areas should not be easily recovered). In contrast, even if the previous methods was able to realize Visibility and Transparency, there are many cases to have that the algorithm is a secret as grounds about the safety, and there is little technique how enough consideration is performed. In this paper, we propose a visible watermarking algorithm which meets the requirements of Visibility, Transparency, and Security. In addition, we confirm the result by an experiment.

1　はじめに

近年，静止画像や映像や音声などのディジタルコンテンツは，品質を劣化させることなく容易にコンテンツの改竄やコピーが可能であるため，ディジタル著作物の著作権保護をどのように行えばよいかが問題になっている．この問題を解決する手段として，電子透かしが一つの手段としてあげられる．電子透かしは大きく分けると，可視型電子透かしと不可視型電子透かしの 2 つがあげられる．後者は，画像に見えない透かし情報を埋め込むことで，第三者は確認することができない．そのため追跡調査や著作権侵害に接触する際に，埋め込まれた透かしを復号して，問題解決の一助に利用するだけであるため，電子透かしの著作権保護の立場から見解すると消極的な手段である．一方で前者は，コ

ンテンツの再利用を防止したい人が，透かし情報としてロゴマークなどを画像全体または価値のある部分に人間が見える形で表示することによって，不正利用者に心理的な圧力・抑制を与えることができる．さらに会社のロゴマークなどを埋め込むことで宣伝効果も期待できる．しかしながら，コンテンツに可視型電子透かしを入れたままにすると，正当なユーザによるコンテンツの再利用もできなくなる．よって，可視型透かしには正当なユーザは透かしを除去できるようにしたいという要求がある．また，コンテンツを完全に隠すと画像の内容自体がわからなくなる．そのため，透かしにある程度の透明性を持たせ，原画像を確認したいという要求も可視透かしには存在する．さらに，可視型電子透かしを入れても正当なユーザ以外に簡単に削除されてしまっては著作権保護の手段としては意味がない．よって除去に関する情報を知らなければ，可視型透かしの除去が困難であるという要求もある．以上から，一般的に可視型電子透かしには以下の 3 つの要件 (a) があると考えられる．

可逆性 透かし情報を埋め込んだ画像から完全に原画像を復元できる

透明性 透かし情報を埋め込んでも原画像をある程度確認できる

安全性 ある情報を知らなければ透かし情報の削除・修復が困難である

可視型電子透かしに関する研究は，静止画像を対象としていくつかの手法が提案されている [1-4]．しかし従来の手法は，可視型電子透かしに関する上記要求が満たされていないものがほとんどである．特に，安全性に関してはそのアルゴリズムが秘密であることを安全性の根拠としている場合が多く，十分な考察が行われている手法は少ない．

本論文では可視型電子透かしに対して行われると考えられる攻撃を分類・検討し，共通鍵暗号の一種であるストリーム暗号を用いて擬似乱数を発生させ，法演算を行い，ユーザが任意に選択した範囲内に画素値が収まる方式を用いることで，要件 (a) を満たし静止画像に適した可視型電子透かしの手法を提案する．また，可視型電子透かしに対して今まで提案されている手法をまとめ，その安全性を検討し，従来文献と提案手法の安全性の比較を行う．

2 可視型電子透かしに対する攻撃の分類

一般に，可視型電子透かしのアルゴリズムは公開されていないが，ここでは透かし除去に関連する秘密鍵に関する情報を除き，可視型電子透かしに関するアルゴリズムは公開されているとする．また，透かし情報は目に見えるのでその形状，および透かし情報が埋め込まれている座標などは知られているものと考える．以下に，可視型電子透かしに対する攻撃を分類する．

2.1 画像修復攻撃

透かし情報が埋め込まれた画像に対して，攻撃者は埋め込まれた画像の画素値などを類推して，透かし情報の削除を行い，原画像を復元する攻撃が考えられる．この画像修復攻撃は，様々な攻撃手法が提案されている [5]．image inpainting の攻撃を例にとると，この攻撃手法は透かし情報を埋め込んだ箇所を取り除くための画像修復技術である．これは画像として不要な部分を指定して周辺の画素値から補間することで修復を可能にする．以下，このような攻撃を攻撃 (b) とする．

2.2 可視透かし画像の偽造攻撃

この攻撃の代表として透かし情報の複製攻撃があげられる．これは，同じ鍵を用いて作成された透かし情報を複製し，似た画素値をもつ他の部分に張り付ける攻撃である．例えば,「1」という形状を可視透かしとして埋め込んだ場合，その透かしを複製して近くに張り付ければ「1 1」という形状が可視透かしとして構成される．この攻撃が用いられる例として，正当なユーザが自分の識別番号を「1」として，攻撃者の識

別番号を「１１」と定義されていた場合，透かし情報が最初に可視透かしを入れた著作者の意図しないものであり，可視型透かしへの攻撃として有効である．以下，このような攻撃を攻撃 (c) とする．

2.3 原画像攻撃

この攻撃は透かし画像に対する原画像を有している場合に可能な攻撃である．可視型透かしの埋め込みを行うサイトなどがある場合，攻撃者は自分の画像を原画像としてそのサイトに埋め込みを依頼することでこの攻撃が可能になる．すなわち，自分が得た透かし画像と原画像を解析して，その結果を基に他の画像を攻撃する．この場合，攻撃者は鍵が封印された透かし埋め込み装置 (抽出装置) をブラックボックスとしてもち，原画像を入力すればそれに対応する可視透かし画像受け取ることができる状況であると考えられる．以下，このような攻撃を総称して攻撃 (d) とする．

2.4 上書き攻撃

この攻撃は，攻撃者はある可視透かし画像またはそれを変更した画像を原画像として新たな可視透かし画像を作成する攻撃のことである．以下，このような攻撃を攻撃 (e) とする．

3 従来文献の評価

以下に，文献 [1–4] の従来方式を説明し，前述の可視型電子透かしの要求 (a)，特に安全性を満たしているかを議論する．

3.1 ウィセッストらの手法 [1]

カラー画像固有の性質を利用した可変表示型の透かし情報の埋め込みを提案している．この手法は可視型電子透かしの要求 (a) に対して，透明性は透かし値を変化させることで満たしているが，可逆性は透かし値に依存し値が大きいと満たさない．したがって，可逆性は常に満たしているとは言えない．(b) に対する安全性は，不可視透かしを埋め込むことで安全性を持たせることが可能であるが，画像に応じて適切な透かし値を見つける必要があり，運用性が良くない．(c) に対しては，同じ画素値をもっている画素と入れ替えた場合，透かし値が同じであれば偽造を検出することができない．(d) に対して，透かし情報が既知のため，透かし値を容易に求めることができる．(e) に対して，上書き攻撃が行われた場合の安全性の検討がされておらず，安全性がないと言える．

3.2 沖原らの手法 [2]

2 つのパラメータを用いて原画像の RGB を変化させることで，可視透かし画像を得ることができる．この手法は可視型電子透かしの要求 (a) に対して，2 つのパラメータの値を画像に応じて便宜変更することで透明性のバランスのとれ，かつ可逆性を満たした可視透かし画像が作成できる．しかし，(b)-(e) に対する安全性は考慮がされていないため，安全性がほとんどないと言える．

3.3 Tsai らの手法 [3]

統計的にある性質を満たす乱数を用いることで，可視透かし画像を生成する手法の提案を行っている．この手法は可視型電子透かしの要求 (a) に対して，可逆性・透明性に関して要求を満たしている．安全性に関して (b) は，疑似乱数の値を大きくすることで透明性の低下，影響の高い個所に埋め込むことで安全性を持たせることができるが，文献 [5] の提案手法に対する安全性の関しては考慮されていない．また，(c) の攻撃に対しては，改竄を検出することができない．(d) に対しては，正規分布の値が小さい乱数は安全でないが，正規分布の値を大きくできれば安全性を増すとしている．しかし，この乱数について予測困難性などに関する議論が行われておらず，かつ具体的な生成法も示されていない．また (e) に関しても，上書きされてしまった場合，どちらが正当であるか証明できない．よっ

て，安全性について正しく評価されていないと言える．

3.4 Huang らの手法 [4]

可視型アルゴリズムを用いて著作権保護を行い，可逆データを用いることで原画像の復元を行う手法である．この手法は可視型電子透かし要求 (a) は，可逆データを用いることで可逆性，埋め込みアルゴリズムを用いることで透明性は持っている．しかし，安全性に関しては，考慮されておらず沖原らの手法 [2] とほぼ同じである．

4 3 つの要件を満たす可視型電子透かしの手法

本提案方式は，共通鍵暗号の一種であるストリーム暗号を用いることで，上記の攻撃に対して安全性を持たせることを実現する．本提案方式の特徴は，透明性を実現するために埋め込む透かし情報（乱数値）の範囲を任意に選択可能になっていることである．本提案方式では，AES などの現時点で安全とされているブロック暗号を用いてストリーム暗号用の擬似乱数を発生させる．選択範囲内の乱数と原画像の画素値を加算して，法演算を行うことで，変換後の画素値が指定範囲に収まる．安全性を考慮すると真性乱数の方がよいが，実用性から今回は擬似乱数を用いる．

4.1 透かし情報の埋め込み手法

原画像にグレースケールの透かし情報を埋め込む．ここで，透かし情報を埋め込む箇所は原画像の任意の箇所に設定可能である．以下の提案手法はグレースケール画像 (1 画素 8 ビット構成) の階調数 256 を用いた際の例として説明する．

step0 実現したい透明性を考慮して，選択範囲 $[0 \sim \epsilon](0 \leq \epsilon \leq 255)$ の値を決める．また，可視透かしを埋め込みたい画素 $p_i(0 \leq p_i \leq 255\ ,\ 0 \leq i \leq n)$ の数を n として，シリアル番号を付ける．

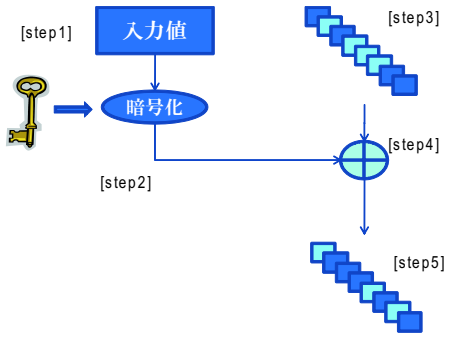


図 1: 透かし情報の埋め込み手法

step1 ブロック暗号を用い，画像の持ち主が作成した秘密鍵を，秘密情報として擬似乱数 r を発生する．ここで，ブロック暗号への入力値として，透かし情報を埋め込む座標と埋め込む透かしの値や画像の識別番号などを用いる．

step2 一般にブロック暗号の出力は 2 値系列であるので，step1 で得られる疑似乱数 r も 2 値系列として得られる．よって，擬似乱数 r を選択範囲内の擬似乱数に変換するため ϵ 進数変換する．すなわち，r を以下の式で表現する

$$r = r_0 + r_1 \times \epsilon + r_2 \times \epsilon^2 + r_3 \times \epsilon^3 \times + \cdots \quad (1)$$

step3 透かし情報を埋め込む画素 p_i の画素値を ϵ 進数変換する．すわなち，

$$\begin{aligned} p_i &= p_{i'} + p_{i_1} \times \epsilon + \cdots + p_{i_m} \times \epsilon^m \\ &= p_{i'} + p_{i''} \times \epsilon \end{aligned} \quad (2)$$

$$(p_{i''} = p_{i_1} + p_{i_2} \times \epsilon + \cdots + p_{i_m} \times \epsilon^{(m-1)})$$

step4 画素 p_i の選択範囲の値をストリーム暗号化し，暗号化画素 q_i を生成する．

$$q_i = (p_{i''} \times \epsilon) + \{(r_i + p_{i'}) \bmod \epsilon\} \quad (3)$$

step5 step4 で生成された値を 2 進数変換し画素に埋め込み，可視透かし画像を生成する．

以上の流れを図 1 に示す．

4.2　透かし情報の抽出手法

step1,2,3 埋め込み処理と同様.

step4 可視透かし画像画素 q_i の選択範囲の値をストリーム暗号化し，原画像画素 p_i を生成する.

$$p_i = (q_{i''} \times \epsilon) + \{(q_{i'} - r_i) \bmod \epsilon\} \quad (4)$$

step5 step4 で生成された値を 2 進数変換し画素に埋め込むことで，原画像が復元できる.

5　実験結果

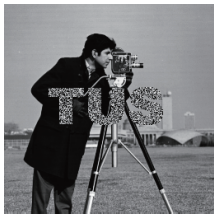
今回の実験では，原画像のサイズ 256 × 256，階調数 256，透かし画像 120 × 58 の画像を用いた．今回は実験画像として lenna と cameraman，透かし情報として [TUS] の文字を用いた．選択範囲は任意設定可能であり，グレースケール画像に ϵ=130,220．カラー画像に ϵ=120 の 3 パターンのパラメータを図 2 に示す．なお，カラー画像はグレースケールで行った提案手法を R,G,B の各々に適応させて行った.

図 2 より ϵ が小さいと透かし情報の透明性が高く原画像に近い画像になる．一方で，ϵ が大きくなると画像の透かし情報がはっきり知覚でき，かつ原画像と無関係のパターン（透明性が低下）となることがわかる.

6　可視型電子透かしの要件 (a) に対する評価

以下では，用いる共通鍵暗号は安全として評価する．また，従来文献の各手法と本提案方式に対する評価結果を表 1 にまとめる．なお，表 1 では，2 章で述べた攻撃に対しての安全性を考慮し，かつ有効的な場合のみ○としている．それ以外は，×としている．なお，文献 [1] は要件 (a) に対して，可逆性を常に満たしていないため△とする．また，(b) に対しては，運用性が良くないため△とする．文献 [3] は (b) に対して，文献 [5] に対する攻撃の安全性を考慮していないため△とする.


cameraman に対して ϵ=130




cameraman に対して ϵ=220


lenna に対して ϵ=120

図 2: パラメータの値を変更した結果の可視透かし画像

6.1　可視型電子透かしに対する要求

5 章の実験結果で述べたように，適切なパラメータを用いることで画像の詳細を保ちつつ透かし情報が埋め込めていることが確認できる．つまり，透明性は満たされている．また，可逆性に関しては，同様の鍵を利用したユーザが同じ擬似乱数を発生させることで，原画像の復元が可能である．安全性に関しては，以下で述べる.

6.2　画像修復攻撃

Image inpainting などの攻撃に対しては，透かし情報を太い線にしたり，原画像の影響の強い箇所に埋め込むことで，修復を困難にすることができる．たとえ修復しても完全に原画像の復元が不可能なため，PSNR などを用いて原画像と比較することで正当な画像か判断することができる．また，提案方式は選択範囲を任意に

表 1: 要件 (a) と各々の安全性の比較

	要求 (a)	攻撃			
		(b)	(c)	(d)	(e)
文献 [1]	△	△	×	×	×
文献 [2]	○	×	×	×	×
文献 [3]	○	△	×	×	×
文献 [4]	○	×	×	×	×
提案方式	○	○	○	○	○

選択可能なので，ユーザが画像の必要性に応じて透明性と安全性を考慮した画像を作成する．つまり，選択範囲を広くすれば，透明性は少なくなるが安全性を高められる．つまり，透明性と安全性はトレードオフの関係になっている．

6.3 可視透かし画像の偽造攻撃

埋め込まれた透かし情報を複製された場合，ブロック暗号への入力値が異なるため，正当なユーザはその部分を除去できず改竄されたとみなすことができるため，安全性があると言える．

6.4 原画像攻撃

本提案方式では，ストリーム暗号を用いているため，暗号化した部分すなわち可視型透かしとなっている部分は原画像と透かし画像があったとしても，そのストリーム暗号が安全であれば秘密鍵の解析はできない．

6.5 上書き攻撃

上書き攻撃に対しては可視型透かしが挿入された画像は誰でも透かし画像と解析できるので，その場合透かしの挿入をしないようにすればよい．よって，上書き攻撃をするためには透かしを除去した画像を原画像とする必要があるが，6.2 から画像の修復は困難であるので，安全であると言える．

7 まとめ

本提案方式は従来の文献と比べて，可視型電子透かしに対する要件をすべて満たすことができた．この理由として，現時点で安全とされているブロック暗号を用いて擬似乱数を発生させることで安全性が高まり，埋め込みと同様の擬似乱数から透かし情報の抽出を行うことで画像に依存しない原画像の復元が可能である．また，法演算を用いることで画素値が指定した範囲に収まり透明性を満たすことが出来る．今後は，fragile 透かしや不可視の可逆性透かしと組み合わせることで多様な用途に使える透かしを検討していく予定である．

参考文献

[1] ウィセット ピヤピスト，松井 甲子雄，“カラー画像への可変表示型電子透かしの提案”, 情報処理学会論文誌, vol.40, no.12, pp. 4370–4377, Dec 1999.

[2] 沖原 建一, 木下 宏揚, “カラーロゴマークに適した可視型電子透かし”, 2005 年電子情報通信学会基礎・境界ソサイエティ大会, A-4-38, p.107, 2005.

[3] H.C. Huang, T.W. Chen, “Copyright Protection and Annotation with Reversible Data Hiding and Adaptive Visible Watermarking”, ICICIC, vol.2, no.5, pp.292–295, 2007.

[4] H.M. Tsai and L.W. Chang, “A high reversible visible watermarking scheme”, in Proc. Int. Conf. Multimedia and Expo, vol.2, no.5, pp.2106–2109, 2007.

[5] C.H. Huang and J. Huang, “Attacking Visible Watermarking Schemes”, IEEE Trans. Multimedia, vol. 6, no.1, pp.16–30, Feb 2004.