Vol.2010-CSEC-51 No.2
2010/12/10

電子透かしに平均値攻撃耐性を持たせるための一手法

久世慎吾[†]　　岩村惠市[††]

結託攻撃のひとつとして多くのユーザーが結託していくつかの透かし画像を平均化することで、透かし情報などを打ち消す平均値攻撃がある。画像圧縮やノイズ付加などの通常の電子透かし画像に対する攻撃は画質を劣化させるが、平均値攻撃は多くの画像が集まれば集まるほど、平均化された画像を原画像に近づけ画質を向上させる。よって、平均値攻撃は電子透かしにとって最も脅威とすべき攻撃と考えられるが、多くの結託者に対して有効な手法は皆無と言える。そこで、平均値攻撃に対抗する一手法を提案する。

A method to have a robustness against average value attack to digital watermark

SHINGO KUSE[†]　KEIICHII IWAMURA[††]

As one of attacks to watermarking images, There is average value attack which cancel watermark information to average some watermark images in collusion with many users. The attack to usual digital waermark images of the image compression and the noise addition, etc. deteriorate the image quality, but the more many images gather, the more the averaged image improve the image quality as a original image in average value attack. Therefore, average value attack is thought to be an attack that should be threaten the most for digital watermarks, but it can be said that an effective method is nothing for many people which collude with each other. So, it proposes a method that opposes average value attack.

1. はじめに [*]

近年、静止画像や映像や音声などをディジタルコンテンツとして扱うことが日常的となっている。この背景には、パソコンの普及、コンピュータの高性能化に伴うインターネット利用人口の増加がある。しかし、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、ディジタルコンテンツ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。例えば、インターネットを利用するとディジタルコンテンツの発送が容易に出来る利点が得られる反面、その受け取ったディジタルコンテンツの著作権者の許諾を得ない再配布が行われ、不正利用がなされるといった問題である。

この問題を解決する手段として、ライセンス認証、データの暗号化、電子透かしなどが挙げられる。それぞれ、一長一短の手法であるが、本論文では電子透かしに着目する。そもそも電子透かしとは、ディジタルコンテンツの冗長部分に人間が知覚出来ないようにサブ情報を埋め込む技術のことである。著作権の保護のために電子透かしを用いる場合、サブ情報として著作者の署名情報や購入者の利用情報（以下、透かし情報）などをディジタルコンテンツに埋め込む。これによって、不正利用されたディジタルコンテンツから抽出された透かし情報から著作者の認知または不正を行った購入者の特定などができ、ディジタル著作物が不正利用されたことを立証することが可能になる。

一般に、電子透かしに求められる主な要件はコンテンツ自体に埋め込むこと、コンテンツ自体を劣化させないこと、コンテンツの編集・加工、および悪意ある攻撃に対して耐性があること、多くの情報量が埋め込めることとなっており[1]、電子透かしの利用において、このバランスが重要である。しかしながら、著作権保護を目的とした電子透かしに対する最大の要件として、不正利用者の悪意のある攻撃によって、埋め込まれた透かし情報を除去されないことが挙げられる。電子透かしへの攻撃法は、（１）単一画像攻撃（1 枚の電子透かし画像を用いる攻撃）と（２）複数画像攻撃（２枚以上の透かし入り画像や、原画像と透かし入り画像の組み合わせ等の複数の画像を用いて行う攻撃）に大別できる。単一画像攻撃としては、画像処理的な攻撃が一般的であり幾何学的変換攻撃、非幾何学的変換攻撃に大別できる。また、複数画像攻撃としては結託攻撃[2]が代表的である。幾何学的変換とは、画像の拡大縮小や回転、ひずみなど各画素の位置関係が変化する変換のことである。一方で、非幾何学的変換とは、JPEG 圧縮や色差やヒストグラムの変化など、位置関係は変わらないが輝度や色差、彩度、

* [†] 東京理科大学大学院
Tokyo University of Science Graduate school
[††] 東京理科大学
Tokyo University of Science

Vol.2010-CSEC-51 No.2
2010/12/10

画素値を変化させる変換のことである。さらに、結託攻撃とは前記利用者情報などの利用者追跡用情報を消す際に有効な攻撃方法である。利用者情報はコンテンツの利用者ごとに異なるため、購入者間の結託によりお互いのコンテンツを比較することで透かしの埋め込み場所を推定することが可能である。

近年、結託攻撃に対抗しては結託耐性符号[3]が研究されており、結託者の数が少ない場合に有効な符号[4]が提案されている。しかし、多くのユーザーが結託していくつかの元画像を平均化することで、元の透かし情報などを打ち消し、または変更し画像を悪用するという平均値攻撃に対しては、結託者が少ない場合に有効ないくつかの電子透かし方式[5]を除いて有効な手段は提案されていない。この攻撃は結託する人数が多ければ多いほど、埋め込まれた透かし情報をきれいに除去することが可能であり、電子透かしにとって最も脅威とすべき攻撃と考えられる。現在、多くの結託者による平均値攻撃に対して有効な手法は皆無と言える。

ここでは、平均値攻撃を以下のように考える。例えば、原画像を G、各購入者 i 毎に埋め込む利用者情報を Ai とすると、 i への透かし画像は Wi＝G＋Ai となる。
n 人の購入者が平均値攻撃を行うと、得られる画像は

$$\sum_{i=0}^{n-1} W_i / n = (n \cdot G + \sum_{i=0}^{n-1} A_i) / n = G + \sum_{i=0}^{n-1} A_i / n$$ となる。

一般に Ai はユーザー毎に定まるランダムな値であるので、その平均値は 0 になるため、n が十分大きければ

$$\sum_{i=0}^{n-1} W_i / n = G$$ となり、透かし情報は除去される。

そこで、本論文のアプローチとして購入者毎に原画像も微小に変化させることにより、平均値攻撃に耐性を持たせることをを考える。すなわち、購入者 i への原画像を Gi＝G0＋gi として、G0 を原画像の特徴を表すコア部、gi を購入者毎の微小な変更部とする。例えば、G0 を原画像の低周波成分、gi をその高周波成分とすることができる。よって、購入者 i への透かし画像 Wi は Wi＝Gi＋Ai となり、その平均値攻撃の結果は

$$\sum_{i=0}^{n-1} W_i / n = (\sum_{i=0}^{n-1} G_i + \sum_{i=0}^{n-1} A_i) / n = (n \cdot G0 + \sum_{i=0}^{n-1} gi + \sum_{i=0}^{n-1} A_i) / n = G0 + \sum_{i=0}^{n-1} gi / n$$ となる。

ここで、$\sum_{i=0}^{n-1} gi / n = 0$ であれば、$\sum_{i=0}^{n-1} W_i / n = G0$ となり、G より画質の劣る低周波成分のみの画像になる。また、$\sum_{i=0}^{n-1} gi / n$ がノイズ成分のようになっていれば、G0 にノイズを加えた画像となる。よって、このアプローチにより、攻撃者は劣化した画像しか得ることができないことになる。このアプローチを実現するためには以下の要件が必要となる。

① 購入者毎の原画像 Gi＝G0＋gi は原画像 G と比べて劣化が少ない。

② Gi の平均を取った $\sum_{i=0}^{n-1} G_i = G0 + \sum_{i=0}^{n-1} gi / n$ は原画像 G に比べて劣化が大きい。

③ n を増せば増すほど劣化が大きくなる。

よって、上記要件を満足する手法を検討することが本論文の目的となる。

本論文では上記アプローチを実現するために、画像攻撃ツールである Stirmark[6][7] の smallrandomdistortions という攻撃を利用する。Smallrandomdistortions で処理をしておいた数枚の画像を用意し、この処理画像の段階ではほとんど劣化しておらず、平均値攻撃を与えた後には劣化している状態を理想とする。これが実現できていれば十分な平均値攻撃耐性を有しているといえる。

以下、第 2 章において Stirmark を用いた smallrandomdistortions の概要を説明する。第 3 章で今回実装した手法の概要と結果を示し、第 4 章で実装手法の評価、考察をする。

2. Stirmark を用いた smallrandomdistortions

Stirmark は 1998 年に Cambridge 大学で開発されたソフトで，2010 年 11 月現在，第 4 版が最新版である． Stirmark は幾何学的改変、非幾何学的改変などの画像処理を行うソフトで，パラメータを細かく設定できるという特徴をもつ。これらは全て単一画像攻撃（１枚の透かし入り画像を用いる攻撃）であり，複数画像攻撃（２枚以上の透かし入り画像や、原画像と透かし入り画像の組み合わせ等の複数の画像を用いて行う攻撃）の機能は無い． また，Stirmark はコマンドラインベースのプログラムであり，攻撃の出力画像を見る為には別に画像エディタを開く必要がある．

この Stirmark の持つ画像処理機能のひとつに smallrandomdistortions というものがある。

まず図 4 に Stirmark の smallrandomdistortions の実施例として微小な幾何学的歪みを与えた場合の歪ませ方を示す。

ここでは、画像を延ばす、シフト、回転させるといった動作[8]を組み合わせることで歪みを与えている。

Vol.2010-CSEC-51 No.2
2010/12/10

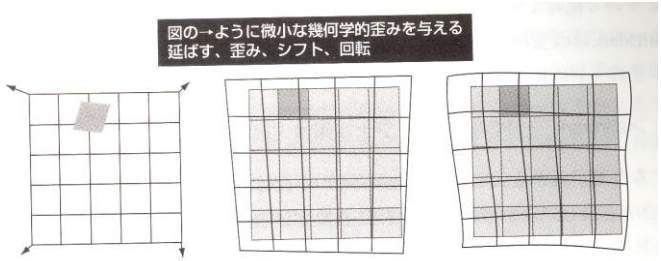


図 1　Stirmark 攻撃の例（画像を歪ませる）

図 1 は文献[8]から出典であるが、図 1 のように画像毎に幾何学的歪みを一定量ではなくランダムに与える操作のことを smallrandomdistortions といい、これは幾何学的改変にさらにランダムな画像歪みが入ることになるので大変強力な攻撃方法である。

3.　提案手法の説明と実験結果

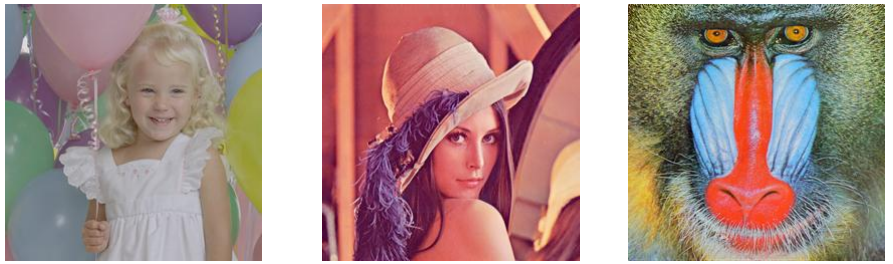
この章では、従来では耐性評価における攻撃手段としてしか用いられていなかった Stirmark を用いて平均値攻撃への耐性を実現させることを考える。この Stirmark の smallrandomdistortions という攻撃では様々な幾何学的変換をさせることができるが、その中でも画像に歪みを与える操作のパラメータを適切に設定することで目視では確認できないほどの歪みを与える事が出来る。つまり、ここでは原画像 G の歪みを与えられていない部分が G0、歪みを与えられた部分が gi となり、その和である歪みを与えられた画像全体が Gi という関係になって Gi=G0+gi となる。これより、1 章のアプローチにおける①の要件は既に満たしていると言える。その処理を施した画像に対して複数の者(n 人)の平均値攻撃があったとき、攻撃後の画像は

$\sum_{i=0}^{n-1} G_i = G0 + \sum_{i=0}^{n-1} gi/n$ となるが、この Gi の平均が原画像 G と比べて劣化が大きければ②の要件を満たすことができる。そして本実験では画像を足し合わせる枚数のパターンをいくつか用いるが、この枚数が多いときに劣化が大きければ③の要件も満たすことになる。以上の①～③の要件を全て満たしているかどうかを評価していく。

3.1　Stirmark を用いる手法

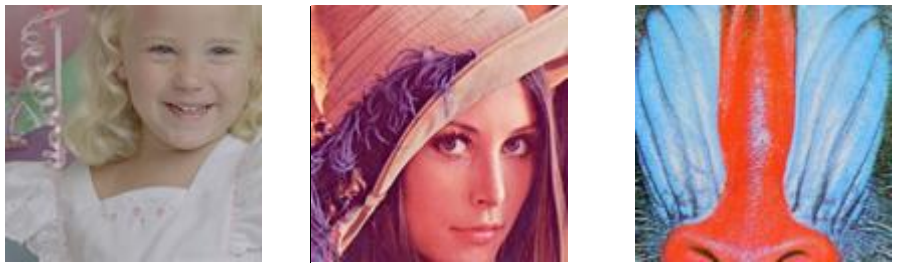
この実験では Stirmark を用いて、画像に対して smallrandomdistortions の攻撃をして画像を歪ませる。そして適当にパラメータの値を決めることで画像の歪みの大きさを少しずつ変えた画像を数枚用意する。これらの画像に平均値攻撃をしたときの耐性をみる。本実験では平均値攻撃において足し合わせる枚数を様々なパターンにして、枚数の違いによる劣化の具合の変化を確かめた。前提条件として図 2 に示すサイズ 256×256 画素、RGB256 階調の 3 種類の BMP のカラー画像を使用する。そして微小な歪みを加えた画像を数名のユーザーに配るものとし、足し合わせる歪ませた画像の枚数分が視覚的に原画像と比較して劣化していない状態であるものを用いる。よって枚数を増やすほど、視覚的に劣化が目立ちやすくなるので画像の中心部を 128×128 画素にトリミングすることでその問題の解決を図った。トリミングした標準画像を図 3 として示す。そして、ユーザーに配る微小な歪みを加えた画像例((a)Balloon の画像を用い、パラメータの値 : 0.01,0.05,0.09,0.125,0.1,0.5,0.9,0.125 の順番とする)を図 4～11 に示す。

図 3 及び図 4～図 11 の結果から 1 枚ごとには画像の歪みの検知は難しいことがわかる。



(a)Balloon　(b)Lenna　(c)Mandrill

図 2　実験に用いた標準画像(256×256 画素,BMP)



(a)Ballon　(b)Lenna　(c)Mandrill

図 3　実験に用いたトリミング画像(128×128 画素,BMP)


図 4　パラメータ 0.01


図 5　パラメータ 0.05


図 6　パラメータ 0.09


図 7　パラメータ 0.125


図 8　パラメータ 0.1


図 9 パラメータ 0.5


図 10 パラメータ 0.9


図 11　パラメータ 1.25

3.2　実験結果

smallrandomditortions を図 2 の標準画像に対して行い、それを神谷が作成した攻撃ツール[7]の平均化攻撃によって数枚を足しこんで平均化したものを 128×128 画素にトリミングした画像を図 3 の画像と比較する。確認方法としては視覚で見ることと、画像の信号と混入したノイズの比率である PSNR の値を見ることで判断する。足しこんだ枚数は 4 枚、8 枚、16 枚、24 枚、48 枚の 5 パターンの実験を行った。

smallrandomdistortions において設定したパラメータは 8 枚以降のパターンは 0.01~0.125 までを 0.005 間隔ずつで 24 枚、0.1~1.25 までを 0.05 間隔ずつで 24 枚の計 48 枚の処理画像を用いた。4 枚のパターンについてはパラメータを別に設定して、行ったが、枚数の少ないときの平均値攻撃に対しては画像の劣化が分かりにくいので、PSNR の値を載せたグラフを図 12 に示す。8 枚、16 枚、24 枚のパターンの(a)Balloon の画像についての比較画像をそれぞれ図 13、図 14、図 15 に示す。3 枚全ての画像についての 48 枚のパターンの比較画像を図 16～18 に示す。右が原画像で、左が平均値画像となっている。

①　4 枚

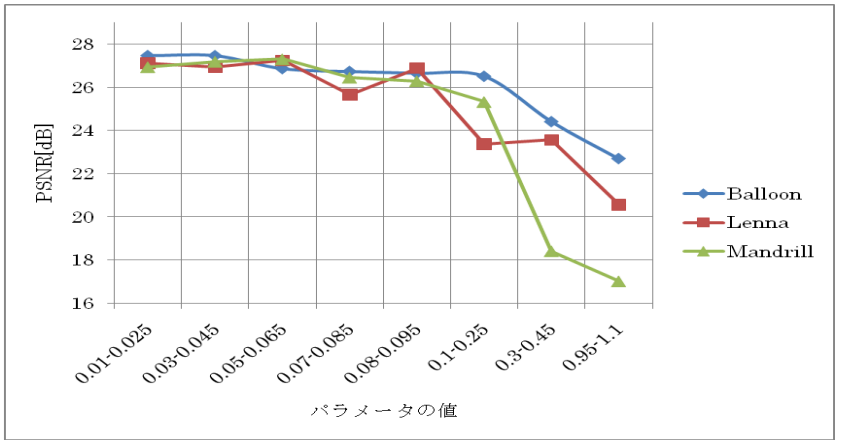


図 12　平均値画像のパラメータの値と PSNR の関係

② 8 枚

図 13 パラメータ 0.01-0.045 の平均値画像と原画像の比較

③ 16 枚

図 14 パラメータ 0.05-0.125 の平均値画像と原画像の比較

④ 24 枚

図 15 パラメータ 0.1-1.25 の平均値画像と原画像の比較

⑤ 48 枚

図 16 パラメータ 0.01-1.25 の平均値画像と原画像の比較(Ballon)

Vol.2010-CSEC-51 No.2
2010/12/10

図 17　パラメータ 0.01-1.25 の平均値画像と原画像の比較(Lenna)

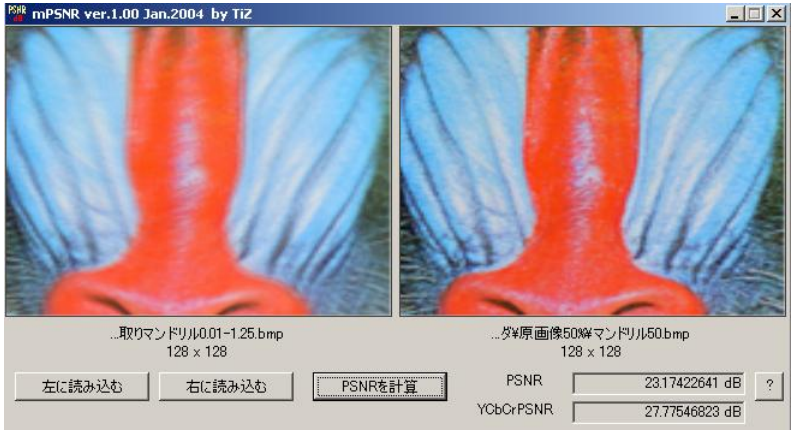


図 18　パラメータ 0.01-1.25 の平均値画像と原画像の比較(Mandrill)

4.　実験評価および考察

4.1　提案手法の実験結果および考察

今回の実験では枚数を足し合わせて平均化して得られた結果は、図 13~図 15 から基本的にはパラメータの値を大きくするほど、そして足す枚数を増やすほど、多少の誤差範囲の上下はするものの劣化が大きくなった。しかし、視覚的には若干の劣化はするのだが、よく確認しなければわからないほどの画像も多かった。ただし、パラメータの値を大きくとった場合は足し合わせる枚数が多くなくとも劣化していることが明らかにわかる。これはパラメータの値が大きいほど原画像に与える歪みの向きや縮尺の変わり方が等間隔の値であっても激しいからだと思われる。ただし、図 16~図 18 をみると枚数を最も多く足し合わせてるので劣化が著しくなってはいるが、Ballon,Lenna に比べると Mandrill の劣化具合は少ない。よって、画像の特徴によっても劣化の仕方が変わってくると考えられ、Mandrill の画像は同じ色である緑、赤等の色が占める割合が多いので同じ面積内において様々な色を使用した Ballon,Lenna よりも歪みを加えても動いた部分が少なくなり劣化が抑えられたと考えられる。

4.2　提案手法と組み合わせる電子透かしの要件についての考察

提案手法と電子透かしを組み合わせるときの電子透かしの要件について考察する。まず提案手法は微小な歪みを加えた後に平均化されるという流れになっている。そこで透かしを埋め込む場合、埋め込むタイミングとして微小な歪みを加える前と微小な歪みを加えた後に埋め込む２パターンが考えられる。また、抽出に関しては原画像が必要な場合と必要でない場合の 2 パターンが考えられる。

まず埋め込むタイミングを微小な歪みを加える前とした場合、一般的な電子透かしを埋め込んだ後に微小な歪みが加わるので、この微小歪みは電子透かしに対する攻撃となる。よって、原画像を用いない抽出を行う場合、用いる電子透かしは smallrandomdistortions のような画像の拡縮、シフト、回転の組合せ攻撃に耐性をもつ手法であることが求められる。また、抽出に際して原画像を用いる場合、原画像を利用して透かし画像の微小歪みを補正するなどの処理を加えれば、よりよい抽出が行える可能性がある。ただし、[5]のように透かし画像と原画像の差分が透かし情報のみになるという前提をもつ場合、補正をしてもその差分画像には比較的大きなノイズが挿入されることが予想されるので、そのままでは利用できない可能性がある。

次に埋め込むタイミングを微小な歪みを加える後とした場合、微小な歪みを加えた後に電子透かしを埋め込むので、この微小歪みは電子透かしへの影響を及ぼさない。よって、原画像を用いない抽出を行う場合、正常な抽出処理が行えると考えられる。しかし、抽出に際して原画像を用いる場合、微小な歪みを加えた画像がそれぞれ原画像となり、透かし画像毎に原画像が異なることになるので多くの原画像を保存しておかなければならず、保存容量的に実現が難しい場合が考えられる。

以上のことから提案手法と組み合わせる透かしの要件は以下のようなものが考え

Vol.2010-CSEC-51 No.2
2010/12/10

られる。

①1 埋め込むタイミングを微小な歪みを加える前とし、原画像を利用せず画像の圧縮、シフト、回転の組み合わせに耐性を持つ

②1 埋め込むタイミングを微小な歪みを加える前とし、原画像を利用して透かし抽出の際透かし画像の微小歪みを適切に補正する手段をもつ

③1 埋め込むタイミングを微小な歪みを加える後とし、原画像は用いない

以上の要件をもつ電子透かしは種々のものが提案されており、本提案は実際に適用することが可能と考えられる。

5. おわりに

本論文では smallrandomdistortions を用いた方法で平均値攻撃に対する耐性評価をした。目的通り、足し合わせる前の画像を劣化させずに平均化したのちに劣化させるという結果を出すことができ、枚数が増えるほど劣化も大きくなることも確認できた。よって、1 章におけるアプローチの要件全てを満たすことができた。

今後は smallrandomdistortions を用いた方法において枚数が多いほど劣化することはわかったが、本実験よりもさらに多い枚数で足し合わせたときどの程度まで劣化するかを確認して耐性を強める方法として使用できるかを試してみたいと思う。また、今後の課題としては実際の透かしとの組合せ、及び提案手法に適した透かし手法の検討なども挙げられる。

謝辞 本研究を進めるにあたって、御指導を頂きました岩村先生に心から感謝致します。

参考文献

1) 社団法人　電子情報技術産業協会,「電子透かし技術に関する調査報告書」2001 年 3 月
2) 浜福　論吉,「入力画像変動を有す電子透かし方式に対する結託攻撃の耐性評価」
3) 磯谷　泰知,　村谷　博文　「結託耐性符号の実用化に向けた符号長の短縮」
4) M. Wu, W. Trappe, Z.J. Wang, and K.J.R. Liu, "Collusion-resistant fingerprinting for multimedia," IEEE Signal Processing Magazine, vol. 21, pp. 15-27, Mar. 2004
5) 加藤寛史, 林直樹, 栗林稔, 森井昌克, "CDMA 技術に基づく電子指紋方式の階層構造の拡大," 2008 年暗号と情報セキュリティシンポジウム (SCIS2008), 2008 年 1 月..
6) K.Kamiya, T.Mori, and K.Iwamura:"Development of Benchmark Tool for Digital Watermarking," 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIHMSP-2008-IS05-009, Harbin, 2008-8.
7) 「Stirmark benchmark」
http://www.petitcolas.net/fabien/ watermarking/Stirmark/
8) 小野　束,「電子透かしとコンテンツ保護」オーム社　2001 年
9) 久永隆治, 栗林稔, 田中初一, "幾何学的歪みの局所的な補正による電子透かし検出能力の改善," 電子情報通信学会論文誌, vol.J88-A, no.10, pp.1146-1153, 2005.