USUWANIE CYFROWYCH ZNAKÓW WODNYCH NA PODSTAWIE CECH AUTOKORELACJI OBRAZU

Dariusz Bogumił
Politechnika Warszawska, Instytut Informatyki
ul. Nowowiejska 15/19, 00-665 Warszawa
dbogumil@elka.pw.edu.pl, http://www.ii.pw.edu.pl/~dbogumil

1. WPROWADZENIE

Gwałtowny rozwój technik informacyjnych i internetu umożliwił niezwykle łatwe kopiowanie i rozpowszechnianie cyfrowych danych. Jednakże w wielu sytuacjach niezbędna jest ochrona praw autorskich lub własności. Ryzyko kradzieży cyfrowego dzieła jest bardzo duże w sytuacji, gdy oryginał i każda kopia są identyczne. Rozwiązaniem może być w takiej sytuacji stosowanie cyfrowych znaków wodnych. Taki znak wodny to niewidoczna informacja zapisana w cyfrowym medium.

Pierwsze znaki wodne pojawiły się około 700 lat temu i służyły jako dowód pochodzenia ręcznie czerpanego papieru od określonego rzemieślnika, jednocześnie potwierdzały typ, format i jakość papieru.

Niemal równolegle z wprowadzeniem znaków wodnych, pojawiły się próby ich fałszowania, oparte zwykle na procesach chemicznych. Mimo to znaki wodne były traktowane jako wiarygodne potwierdzenie autentyczności, czego wyrazem było używanie ich jako materiałów dowodowych przed sądem.

W wieku danych cyfrowych pojawił się podobny problem jak 700 lat temu. O ile zwykłe znaki wodne zabezpieczały i identyfikowały papier, na którym informacje były drukowane lub zapisywane, cyfrowe znaki wodne muszą być oznaczeniem samych danych, ponieważ technika cyfrowa uniezależniła dane od nośnika.

W ciągu ostatnich 10 lat prowadzono intensywne badania w tej dziedzinie. Opracowano wiele algorytmów, technik i działających systemów służących do podpisywania danych cyfrowych. Powstało także kilka efektywnych, komercyjnych rozwiązań. Współczesne znaki wodne mogą służyć zarówno do ochrony praw autorskich czy monitorowania dystrybucji danych (np. w połączeniu z szyfrowaniem umożliwia to wykrycie, kto opublikował zastrzeżone informacje), jak i do zapewniania integralności danych oraz wskazywania zmienionych miejsc w obrazie. Znak wodny może być dowodem tak autorstwa, jak i legalnego zakupu.

Nowy pomysł na zabezpieczanie cyfrowych danych pobudził nie tylko badania nad nowymi technikami tworzenia cyfrowych znaków wodnych, ale także poszukiwania metod usuwania czy podrabiania znaków wodnych, często w celu ominięcia prawa i pobierania nielegalnych korzyści. Pozytywną stroną takich badań jest wskazywanie słabości istniejących systemów, motywujące do szukania nowych, lepszych metod, odpornych na ataki.

Niniejsza praca ujawnia wrażliwość niektórych systemów cyfrowych znaków wodnych na atak bazujący na analizie wartości funkcji autokorelacji obrazu. Dokument przedstawia nową metodę usuwania znaków wodnych z obrazów cyfrowych. Atak bazuje na cechach funkcji autokorelacji zabezpieczonego obrazu i z tego względu dotyczy tych systemów cyfrowych znaków wodnych, w których znak wodny ma ustalone właściwości funkcji autokorelacji. W szczególności obejmuje to szeroko rozpowszechnione systemy dzielące obraz na bloki i wprowadzające kopie znaku wodnego do kolejnych bloków. Nowa metoda umożliwia wykrycie i usunięcie znaku wodnego tylko na podstawie analizy obrazu ze znakiem wodnym. Jednocześnie, w przeciwieństwie do innych znanych ataków, podczas tej operacji wysoce prawdopodobne jest poprawienie jakości przetwarzanego obrazu, rozumianej jako podobieństwo do obrazu oryginalnego.

Rozdział 2 prezentuje dlaczego i w jaki sposób systemy cyfrowych znaków wodnych wykorzystują właściwości autokorelacji. Opis zaproponowanego algorytmu usuwania tak wstawionego znaku wodnego zawarty jest w rozdziale 3, natomiast uzyskane wyniki – w 4.

2. ZNAK WODNY A AUTOKORELACJA OBRAZU

Funkcja autokorelacji jest zdefiniowana jako korelacja obrazu z tym samym obrazem:

 # 2-1
$$R_{u,u}(n,m) = \sum_x \sum_y u(x,y) \cdot u(x+n, y+m)$$

Przydatność analizy wartości funkcji autokorelacji w systemach cyfrowych znaków wodnych spowodowana jest tym, że funkcja ta przechodzi takie same transformacje geometryczne, jak zaatakowany obraz oryginalny. Można to wytłumaczyć tym, że relacje między punktami po przeprowadzeniu liniowej transformacji geometrycznej nie zmieniają się [5]. Na przykład odległość między dwoma punktami przechodzi zmiany proporcjonalne do przekształceń całego obrazu, niezależnie od położenia punktów. Położenie czy przesunięcie obrazu nie ma żadnego wpływu na względne relacje między punktami w obrazie oryginalnym i po przekształceniach.

Problem odczytania znaku wodnego w przekształconym obrazie jest dość trudny, a rozwiązania są mało odporne na inne zakłócenia [2][3][4] lub bardzo czasochłonne [1]. Analiza autokorelacji pozwala na szybkie zidentyfikowanie przekształceń obrazu. Oczywiście różne obrazy nie mają stałej, znanej z góry postaci autokorelacji. Z tego powodu koder musi wprowadzić pewne znane właściwości autokorelacji do obrazu. Modyfikacja taka powinna być niezauważalna dla oka i być odporna na próby usunięcia, czyli powinna mieć podobne cechy jak sam znak wodny. Co więcej, postać funkcji autokorelacji powinna być taka, aby możliwe było łatwe określenie, w jaki sposób został przekształcony obraz. Najprościej jest osiągnąć ten warunek, gdy wykres autokorelacji ma kilka wierzchołków (wówczas można łatwo znaleźć macierz przekształcenia odwrotnego). Efekt taki wystąpi, gdy do obrazu wstawimy kilka takich samych „znaków wodnych", różniących się tylko przesunięciem [6]. Metoda ta jest stosowana w komercyjnych systemach cyfrowych znaków wodnych, np. Digimarc [7], Kodak [8]. Zalety rozwiązania polegającego na wielokrotnym wstawieniu kopii znaku wodnego w kolejnych fragmentach obrazu, to:

- zwiększenie odporności znaku wodnego na przypadkowe usunięcie poprzez np. kompresję JPEG, dzięki nadmiarowości umożliwiającej wykrywanie i poprawianie błędów odczytu
- odporność na kadrowanie (obcinanie) obrazu – znak wodny można odczytać z fragmentu obrazu
- odporność znaku wodnego na liniowe przekształcenia geometryczne obrazu – analiza „wierzchołków" na wykresie autokorelacji umożliwia znalezienie macierzy przekształcenia i zastosowanie operacji odwrotnej.

Okazuje się, że możliwe jest usunięcie znaku wodnego wstawionego tą metodą w taki sposób, że jakość obrazu nie tylko się nie pogorszy, ale co więcej – obraz po ataku będzie bardziej zbliżony do obrazu oryginalnego niż obraz ze znakiem wodnym.

Możliwość takiego ataku została przedstawiona w artykule O. Escoda [9]. W przedstawionym tam algorytmie zakłada się, że znak wodny został wstawiony w równej mierze w obszarach jednolitych i w silnie teksturowanych. Oznacza to, że jeśli obraz jest dostateczne duży, aby cała kopia znaku wodnego mieściła się w jednolitym fragmencie obrazu, można poprzez odjęcie średniej wartości jasności pikseli w tym fragmencie wyodrębnić znak wodny. Następnym krokiem jest odjęcie kolejnych kopii znaku wodnego od obrazu, czego wynikiem jest obraz z całkowicie usuniętym znakiem wodnym. Wydaje się jednak, że przyjęte założenia są zbyt ostre – w praktyce systemy cyfrowych znaków wodnych uwzględniają cechy układu wzroku człowieka, więc w obszarach jednolitych znak jest wstawiany z bardzo małą siłą [10].

W rozdziale 3 przedstawiona została inna metoda, pozwalająca na automatyczne usunięcie znaku wodnego na podstawie analizy wartości funkcji autokorelacji obrazu.

3. USUWANIE ZNAKU WODNEGO

Istota metody polega na spostrzeżeniu, że wyraźne maksima funkcji autokorelacji są spowodowane wielokrotnym wstawieniem tego samego znaku wodnego w różnych fragmentach obrazu. Oznacza to, że poprzez analizę podobieństwa tych obszarów być może uda się wyodrębnić modyfikacje pikseli spowodowane wstawieniem znaku wodnego. Algorytm składa się z kilku kroków:

1. Odfiltrowanie obrazu-nośnika od przewidywanego znaku wodnego
2. Obliczenie wartości funkcji autokorelacji dla przetworzonego obrazu
3. Znalezienie punktów o maksymalnej wartości autokorelacji
4. Wybranie pikseli, dla których czynnik autokorelacji w punktach o maksymalnej autokorelacji był odpowiednio duży
5. Usunięcie przewidywanego znaku wodnego w wybranych pikselach.

3.1. Odfiltrowanie obrazu-nośnika

W systemie detektora cyfrowych znaków wodnych sam znak wodny można interpretować jako poszukiwany sygnał, natomiast obraz cyfrowy to medium, nośnik znaku wodnego, czyli stanowi tylko zakłócenia przy odbiorze właściwego sygnału. W takiej sytuacji warto zauważyć, że zmniejszenie wpływu zakłóceń może poprawić wydajność detektora [2]. Zneutralizowanie wpływu zakłóceń (np. poprzez odjęcie ich od sygnału wejściowego) byłoby więc bardzo korzystne.

Ze względu na to, że moc sygnału znaku wodnego jest wielokrotnie niższa niż moc sygnału obrazu-nośnika, niezwykle ważne jest odseparowanie przewidywanego znaku wodnego od obrazu-nośnika. Jednakże dekoder nie dysponuje obrazem oryginalnym, więc dokładne oddzielenie obrazu od znaku jest niemożliwe.

Dla większości obrazów jako całości trudno określić z góry właściwości statystyczne, można jednak przypuszczać, że jasność sąsiadujących ze sobą pikseli różni się zwykle o niewielkie wartości. W takim przypadku średnia arytmetyczna wartości jasności pikseli z pewnego niewielkiego obszaru obrazu powinna dobrze charakteryzować cechy obrazu oryginalnego, natomiast być w niewielkim stopniu zależna od wstawionego znaku wodnego. Z tego względu odjęcie średniej wartości jasności w sąsiedztwie piksela od jego wartości jasności powinno poprawić wydajność detektora i zmniejszyć stopę błędów.

Z przeprowadzonych eksperymentów wynika, że najlepsze wyniki daje zastosowanie filtru w kształcie krzyża.

3.2. Obliczenie wartości funkcji autokorelacji

Dwuwymiarowa autokorelacja obrazu obliczana jest ze wzoru # 2-1. Ze względu na dużą złożoność obliczeniową tej formuły stosowany jest równoważny wzór:

 # 3-1 $$R_{u,v}(n,m) = \Im^{-1}\{\overline{\Im\{u\}} \cdot \Im\{v\}\}$$

gdzie $\Im\{u\}$ to przekształcenie Fouriera obrazu *u*. Dla obliczania transformaty Fouriera istnieją efektywne algorytmy, pozwala to więc polepszyć wydajność systemu.

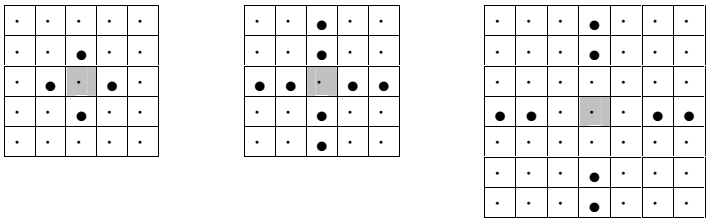
3.3. Znalezienie punktów o maksymalnej wartości autokorelacji

W kroku tym szukane są wszystkie lokalne maksima. Ze względu na niedokładne odfiltrowanie obrazu-nośnika wartości maksymalne mogą się bardzo różnić. Dodatkowo w bezpośrednim sąsiedztwie może wystąpić kilka punktów o dużej wartości autokorelacji – zjawisko niepożądane. Z tego względu najpierw poszukiwane są punkty o największych wartościach autokorelacji, a następnie eliminowane są te punkty, które sąsiadują ze sobą – wybierany jest punkt o wartości największej.

Na listę trafiają te punkty, których wartość autokorelacji jest nie mniejsza niż 0.3 maksymalnej wartości autokorelacji - z wyłączeniem wartości w okolicach punktu (0, 0). Próg został określony w drodze eksperymentów.

3.4. Wybranie pikseli, których wartości się powtarzają

Kroki 3.4 i 3.5 powtarzane są dla trzech różnych kształtów filtrów predykcji, aby możliwie najdokładniej odseparować znak wodny. Zastosowane filtry mają następujące kształty:



Rys. 1

Dla wybranych w poprzednim kroku punktów obliczane jest podobieństwo między pikselami w obrazie nie przesuniętym i przesuniętym o współrzędne kolejnego punktu. Wartość ta obliczana jest dla każdego piksela osobno. Wartość podobieństwa jest inkrementowana, jeśli iloczyn wartości pikseli ma ten sam znak co autokorelacja w punkcie o współrzędnych odpowiadających przesunięciu.

W przypadku, gdy wartość piksela jest większa od określonego progu, piksel ten jest ignorowany w danej fazie obliczeń. Ma to na celu zminimalizowane efektu pogorszenia jakości obrazu po usunięciu znaku wodnego. Próg można konfigurować w programie testowym → parametr „Maksymalne odchylenie wartości piksela”.

3.5. Usunięcie przewidywanego znaku wodnego

Piksele, dla których wartość podobieństwa była odpowiednio duża są modyfikowane. Próg można konfigurować w programie testowym → parametr „Minimalne podobieństwo piksela w powtórzonych fragmentach”.

Modyfikacja polega na odjęciu od piksela w obrazie źródłowym wartości uzyskanej po zastosowaniu odpowiedniego filtru predykcji (jak w punkcie 3.4). Dodatkowo, wartość ta może być pomnożona przez czynnik losowy (zwiększa skuteczność usuwania znaku wodnego), który ma rozkład Gaussa o średniej równej 1 i odchyleniu standardowym

konfigurowanym w programie testowym → parametr „Odchylenie standardowe czynnika losowego".

Aby poprawić subiektywną jakość obrazu postrzeganą przez człowieka, modyfikacje są zmniejszane, jeśli wariancja otoczenia piksela jest większa niż wartość progowa, określona przez parametr „Maksymalna wariancja otoczenia piksela".

4. BADANIA EKSPERYMENTALNE I UZYSKANE WYNIKI

Przedstawiona metoda została zaimplementowana w programie „Cyfrowe znaki wodne" [11]. Badania eksperymentalne polegały na próbach usunięcia znaku wodnego wstawionego za pomocą programu Digimarc (składnik pakietu Photoshop i wielu innych komercyjnych programów do obróbki obrazu).

Zaobserwowano wysoką skuteczność usuwania informacji znaku wodnego dla obrazów w odcieniach szarości. Gorsze wyniki uzyskano dla obrazów w 24-bitowym kolorze. Wynika to stąd, że w zaimplementowanym systemie obróbka pikseli odbywa się tylko w dziedzinie luminancji. Prawdopodobnie działanie niezależnie na każdej składowej koloru pozwoliło by uzyskać wyniki takie jak dla obrazów monochromatycznych.

Okazało się, że możliwe jest usunięcie informacji znaku wodnego lub przynajmniej szablonu do synchronizacji w taki sposób, że obraz po obróbce będzie miał lepszą jakość (rozumianą jako podobieństwo do obrazu oryginalnego) niż obraz ze znakiem wodnym.

Przykładem może być obraz lena.bmp, dla którego PSNR obrazu ze znakiem wynosiło 34,94 dB, natomiast PSNR po usunięciu znaku wodnego 37,30 dB. Jest to dość duża różnica, wyraźnie widoczna także przez subiektywne porównanie badanych obrazów przez człowieka. Wyniki zostały przedstawione w tabeli niżej. Wszystkie badane obrazy miały wielkość 512x512 pikseli w 256 odcieniach szarości. Parametry do usunięcia znaku zostały dobrane eksperymentalnie i oznaczają kolejno:

- minimalne podobieństwo piksela w powtórzonych fragmentach
- maksymalna wariancja otoczenia piksela
- maksymalne odchylenie wartości piksela
- odchylenie standardowe czynnika losowego

Obraz	boat.bmp	baboon.jpg	lena.jpg	lena.jpg
Siła wstawienia znaku wodnego	4	4	4	3
PSNR obrazu ze znakiem wodnym	34,01	32,70	34,94	38,22
Parametry do usunięcia znaku	70-50-20-0	70-175-30-0	85-40-20-0	75-40-15-0
PSNR obrazu z usuniętym znakiem	34,94	30,96	37,30	38,98

Tabela 1

System Digimarc wstawia ten sam znak wodny (lub tylko szablon do synchronizacji – dokumentacja tego systemu nie jest dostępna publicznie) co 128 pikseli. Wynika z tego, że w obrazie o wymiarach 512x512 szablon ten jest wstawiony 16 razy. Dla takich obrazów najlepsze wyniki uzyskano dla wysokiego progu podobieństwa piksela w powtarzających się fragmentach obrazu (ok. 65-90%).

Nawet w przypadkach, gdy znak wodny nie został usunięty całkowicie, zwykle niszczone były skutecznie właściwości autokorelacyjne obrazu. Dzięki temu aby uniemożliwić odczytanie znaku wodnego wystarczy nieznacznie przeskalować obraz, obrócić go czy tylko przesunąć.

5. PODSUMOWANIE

W pracy przedstawiono metodę usuwania znaku wodnego na podstawie analizy wartości funkcji autokorelacji obrazu. Uzyskane w eksperymentach wyniki są bardzo dobre,

co sugeruje, że systemy cyfrowych znaków wodnych, które wykorzystują właściwości autokorelacji obrazu, są narażone na skuteczny atak tego typu. Rozwiązaniem tego problemu może być stosowanie bardziej skomplikowanych algorytmów niż proste wyszukiwanie „wierzchołków" na wykresie funkcji autokorelacji.

Dodatkowo zauważono, że przewidywany znak wodny może być dobrze przybliżony poprzez zastosowanie po kolei kilku filtrów predykcji, zamiast jednego. Możliwie dokładne przybliżenie znaku wodnego ułatwia bezbłędny odczyt ukrytej informacji, z tego względu jest ważne we wstępnej fazie dekodowania znaku wodnego. W programie testowym zastosowano trzy różne rodzaje filtrów. Wybór rodzajów i kolejności stosowania filtrów może być przedmiotem dalszych badań.

6. BIBLIOGRAFIA

[1] F. Hartung, Jonathan K. Su, Bernd Girod. *Spread Spectrum Watermarking: Malicious Attacks and Counter Attacks*. W *Proceedings of SPIE Vol. 3657, Security and Watermarking of Multimedia Contents*, San Jose, styczeń 1999, s. 147-158.
<http://www.nt.e-technik.uni-erlangen.de/LNT_I/publications/pub_list/pub_files/lnt1999_008.pdf>

[2] Joseph J.K. O'Ruanaidh, Thierry Pun. *Rotation, scale and translation invariant spread spectrum digital image watermarking*. W *Signal Processing* 66 (1998), s. 303-317.
<www1.elsevier.nl/cas/tree/store/sigpro/sub/1998/66/3/1170.pdf>

[3] Joseph J.K. O'Ruanaidh, Thierry Pun. *Rotation, Scale and Translation Invariant Digital Image Watermarking*. Centre Universitaire d'Informatique, Universite de Geneve.
<cuiwww.unige.ch/~vision/Publications/postscript/98/ORuanaidhPun_sp98.ps.gz>

[4] Shelby Pereira, Thierry Pun. *Fast Robust Template Matching for Affine Resistant Image Watermarks*. University of Geneva.
<cuiwww.unige.ch/~vision/Publications/postscript/99/PereiraPun_wih99.ps.gz>

[5] Martin Kutter. *Digital image watermarking: hiding information in images*. École Polytechnique Fédérale De Lausanne.
<ltswww.epfl.ch/kutter/watermarking/publications/WIPE.ZIP>

[6] E. Debes, G. Dardier, T. Ebrahimi, A. Herrigel. *Watermarking scheme for large images using parallel processing.* W *In proceedings of SPIE and IS&T conference on Security and Watermarking of Multimedia Contents III*. San Jose, styczeń 2001.

[7] Digimarc
<www.digimarc.com>

[8] Chris Honsinger, Majid Rabbani. *Data Embedding Using Phase Dispersion*. Imaging Science Division, Eastman Kodak Company, Rochester, NY USA
<http://www.kodak.pl/US/plugins/acrobat/en/corp/researchDevelopment/dataEmbedding.pdf>

[9] Oscar Divorra Escoda, Rosa Maria Figueras i Ventura, Eric Debes, Touradj Ebrahimi. *Influence of a Large Image Watermarking Scheme Parallelization on Possible Attacks.* W *Proceedings of the SPIE's Annual Meeting on Optical Science and Technology*, San Diego, California, wrzesień 2001.
<http://dewww.epfl.ch/~debes/pub/SPIEWMAttacks.pdf >

[10] Sviatoslav Voloshynovskiy, Alexander Herrigel, Nazanin Baumgaertner, Thierry Pun. *A Stochastic Approach to Content Adaptive Digital Image Watermarking.* University of Geneva.
<cuiwww.unige.ch/~vision/Publications/postscript/99/VoloshynovskiyHerrigelBaumgaertnerPun_wih99.ps.gz>

[11] Dariusz Bogumił. *Cyfrowe znaki wodne odporne na kompresję JPEG.* Praca dyplomowa, Politechnika Warszawska, wrzesień 2001.
<http://www.ii.pw.edu.pl/~dbogumil>