Metody Informatyki Stosowanej, nr 2/2007
Kwartalnik Komisji Informatyki Polskiej Akademii Nauk Oddział w Gdańsku

Odporność graficznych znaków wodnych na wybrane ataki

Mirosław Łazoryszczak, Piotr Boryszek

Politechnika Szczecińska, Wydział Informatyki

Abstract:

Digital watermarking is one of intellectual properties protection methods. It could protect several media. However the most popular are graphics and sound. This paper shortly presents selected, well known methods of watermarking of the digital images and their level of immunity to the different attacks. First of all these attacks are typical transformations such as lossy compression, different filtrations etc. A combined attack is also introduced. This attack consists of some usual transformation usually used separately. But in the real world the images are often changed many times in many ways.

Słowa kluczowe:

digital watermarking, copyright protection

1. Wstęp

Cyfrowy znak wodny (CZW) jest jednym ze sposobów ochrony własności intelektualnej. Stanowi on ukrytą bądź jawną informację zintegrowaną z materiałem źródłowym. Informacja ta w przypadku graficznych znaków wodnych może być widoczna bądź niewidoczna, zależnie od funkcji, którą ma pełnić w danym medium.

Cyfrowy znak wodny spełnia swą funkcję wówczas, gdy jest zintegrowany z medium w sposób, który nie powoduje widzialnej zmiany sygnału nośnego. Kolejnym wymogiem stawianym CZW jest takie jego umieszczenie w kontenerze, aby jego usunięcie było praktycznie niemożliwe. W praktyce powinno to oznaczać, iż próba usunięcia CZW z sygnału nośnego skutkuje możliwie największym zniekształceniem sygnału źródłowego, czyli takim, który uniemożliwia dalsze, zgodne z normalnym przeznaczeniem wykorzystanie danego materiału graficznego. CZW powinien być także odporny na różnego rodzaju przekształcenia technologiczne jak i celowe modyfikacje. Może zachodzić także potrzeba zachowania wewnętrznej integralności przy podziale pliku zawierającego dany obraz. Wynika z tego, iż CZW powinien być wbudowany w dane środowisko wielokrotnie. CZW nie powinien być także podatny na wielokrotną kompresję oraz dekompresję sygnału źródłowego. Ten ostatni wymóg jest szczególnie ważny w przypadku udostępniania materiału graficznego za pośrednictwem Internetu. Pliki muzyczne pobierane np. w formacie

Metody Informatyki Stosowanej, nr 2/2007
Kwartalnik Komisji Informatyki Polskiej Akademii Nauk Oddział w Gdańsku

JPG zawierającym CZW przekształcane bywają często na format mapy bitowej i mogą być w tej postaci rozprowadzane dalej. Powtórna kompresja do tego samego formatu wejściowego zwykle powoduje istotną degradację znaku wodnego.

W niniejszym artykule przedstawiono porównanie odporności wybranych metod umieszczania znaków wodnych w materiale graficznym na niektóre, często stosowane przekształcenia obiektów graficznych.

2. Umieszczanie znaków wodnych w środowisku graficznym

Wykorzystanie CZW do ochrony plików graficznych możliwe jest dzięki niedoskonałości systemu percepcyjnego człowieka. Dotyczy to przede wszystkim takich zmysłów jak słuch czy wzrok. Podstawową zależnością wykorzystywaną do ukrywania informacji jest prawo Webera-Fechnera (1), które mówi, iż stosunek ledwo dostrzegalnego przyrostu wrażenia do wielkości danego bodźca jest wielkością stałą [2].

$$\Delta I/I = \text{const.} \tag{1}$$

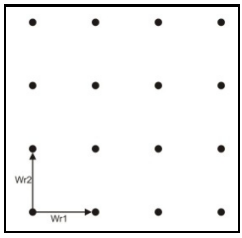
W praktyce możliwość ukrywania informacji bardzo ściśle zależy od zawartości kontenera, którym w tym przypadku jest obraz cyfrowy. Np. szum jest bardzo widoczny na jednolitych powierzchniach takich jak niebo, ulica, a mniej widoczny w pobliżu wszelkich krawędzi oraz ciemnych lub bardzo jasnych powierzchni [2].

3. Wybrane metody znakowania wodnego

Poniżej przedstawiono pięć spośród wielu istniejących metod służących do wbudowania cyfrowych znaków wodnych w obrazach cyfrowych.

3.1. Metoda kodów kratowych (LATTICE)

Metoda ta wykorzystuje ortogonalną przestrzeń kratową (rys. 1). Dzięki niej można zakodować jeden bit na każde 256 pikseli warstwy wybranego koloru. Zatem biorąc obraz kolorowy o rozmiarach 350×350 pikseli (np. „baboon.jpg") mamy do dyspozycji $(350 \times 350 \times 3)/256 = 1435,5$ bitów. Bity znaku są wpierw kodowane z pomocą kodu Trellis, co zwiększa ich odporność na przekłamania oraz powoduje konieczność stosowania w procesie dekodowania algorytmu Viterbiego [4].

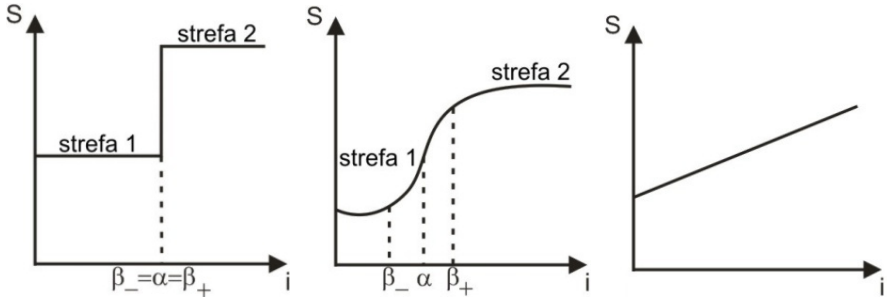


Rysunek 1. Dwuwymiarowa ortogonalna przestrzeń kratowa [4]

Metody Informatyki Stosowanej, nr 2/2007
Kwartalnik Komisji Informatyki Polskiej Akademii Nauk Oddział w Gdańsku

3.2. Metoda Bruyndonckx (BRUYN)

Metoda Bruyndonckx bazuje na klasyfikacji pikseli. Obraz dzielony jest na bloki, którym następnie jest przydzielana jedna z trzech stref (rys. 2) – o silnym (ang. *hard contrast*), progresywnym (ang. *progressive contrast*) lub zróżnicowanym (ang. *noise contrast*) kontraście [1]. Następnie piksele są dzielone na kategorie, w zależności od strefy w której się znajdują. Ich wartość zostaje zmieniona uwzględniając strefę, kategorię oraz wartość bitu informacji.



Rysunek 2. Strefy kontrastu: a) silny, b) progresywny, c) zróżnicowany [Bruyndonckx95]

3.3. Metoda Rosa (ROSA_DCT)

Metoda Rosa oparta jest na dziedzinie częstotliwości [7]. Do zapewnienia możliwości odzyskania informacji posłużono się podejściem zaproponowanym przez Piva, które zakłada, że aby móc udowodnić fakt osadzenia informacji w danym kontenerze bez konieczności posiadania oryginału należy wykorzystać dużą ilość współczynników dyskretnej transformaty kosinusowej (DCT), a także dysponować oryginalną wiadomością [6]. Za pomocą tej metody osadzany jest w obrazie ciąg pseudolosowy. Wyznaczane są współczynniki DCT, które będą przechowywały informację. Dodatkową modyfikacją usprawniającą algorytm było wykorzystanie właściwości systemu percepcyjnego człowieka.

3.4. Metoda z ustaloną percepcją (PERC_GSCALE)

Metoda z ustaloną percepcją to kolejna metoda wykorzystująca model Watsona [4]. Detekcja oznakowanego obrazu jest możliwa dzięki zastosowaniu wybielania liniowego. Polega ono na filtracji obrazu przy pomocy filtru wybielającego (Tabela 1). Obraz dzielony jest na fragmenty 11×11 pikseli, które są następnie przemnażane przez odpowiednie współczynniki filtra. Otrzymane wartości są sumowane, a wynik wstawiany w środek fragmentu nowego obrazu. Analogicznie postępuje się ze wzorcem, a następnie oblicza ich korelację. W ten sposób otrzymujemy wynik – informację czy dany kontener oznakowano testowanym znakiem.

Tabela 1. Współczynniki filtru wybielającego [4]

0.0	0.0	0.0	0.0	0.1	-0.2	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.1	-0.3	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.2	-0.5	0.2	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.1	0.4	-1.1	0.4	0.1	0.0	0.0	0.0
0.1	0.1	0.2	0.4	1.8	-5.3	1.8	0.4	0.2	0.1	0.1
-0.2	-0.3	-0.5	-1.1	-5.3	15.8	-5.3	-1.1	-0.5	-0.3	-0.2
0.1	0.1	0.2	0.4	1.8	-5.3	1.8	0.4	0.2	0.1	0.1
0.0	0.0	0.0	0.1	0.4	-1.1	0.4	0.1	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.2	-0.5	0.2	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.1	-0.3	0.1	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.1	-0.2	0.1	0.0	0.0	0.0	0.0

3.5. Metoda DWT z adaptacyjnym progiem kodowania (KIM)

Metoda zaproponowana Kima bazuje na transformacie falkowej [5]. W procesie detekcji metody potrzebny jest zarówno oryginalny obraz jak i oryginalny cyfrowy znak wodny. Znak jest osadzany w wybranych współczynnikach falkowych w zależności od współczynnika skali (ang. *level-adaptive thresholding*). Współczynnik ten jest zależny od poziomu dekompozycji sygnału. Polega ona na zwiększaniu poziomu szczegółów z każdym kolejny zwiększeniem jej stopnia. W procesie detekcji wyznaczane jest podobieństwo wektora oryginalnego znaku z tym znalezionym w obrazie.

4. Wyniki badań

Procedura testowa zakładała określony schemat działania. Przy starcie następowało ustawienie parametrów startowych wybranej metody. W kolejnych krokach kodowano wiadomości w kontenerze, dokonywano odczytu wartości współczynnika *PSNR* po kodowaniu względem kontenera źródłowego. W przypadku braku akceptacji kontenera wybierano inne parametry i przeprowadzano ponowne kodowanie wiadomości w kontenerze. W przypadku zadowalającej jakości kontenera wybierano moc ataku, przeprowadzano atak (przekształcenie), obliczano wartość współczynnika *PSNR* oraz oceniano skuteczność ataku.

Jako ataki testowe wybrano następujące przekształcenia: kompresja stratna JPEG i JEPG2000 (przy czym z uwagi na różnice w działaniu kompresja JPEG wykazała się znacznie bardziej niszczącym działaniem zarówno w stosunku do oryginału jak i znaku wodnego), atak szumem, wyostrzanie, wygładzanie, usuwanie szumów, zmiana rozmiarów.

Obrazy testowe zawierające znak wodny poddano także atakowi kombinowanemu, składającemu się z następujących przekształceń:

— usuwanie szumów,
— wygładzanie,
— wyostrzanie,

— Kompresja JPEG 2000 w stopniu 85%,
— Kompresja JPEG w stopniu 5%,
— Zmiana rozmiaru obrazu z interpolacją.

a) b) c)

Rysunek 3. Kontenery po ataku kombinowanym: a) baboon.jpg (PSNR=31.7194 dB), b) lena.bmp (PSNR=38.5304 dB), c) peppers.png (PSNR=41.3002 dB)

Na rys. 3 przedstawiono obrazy po ataku kombinowanym. W każdym przypadku otrzymano obraz bez artefaktów, ani innych widocznych zmian. Choć wartość współczynnika PSNR jest dość niska i osiąga wartości poniżej wcześniej ustalonego akceptowalnego progu, należy uznać obrazy jako bardzo dobrej jakości, praktycznie nie odbiegające od obrazu źródłowego.

Rysunek 4. Rezultat odporności wybranych znaków wodnych na działanie ataków(1-BRUYN, 2-LATTICE, 3-ROSA_DCT, 4-PERC_GSCALE, 5-KIM)

Rys. 4 ilustruje otrzymane wyniki testów odporności wybranych metod znakowania wodnego na najczęściej występujące przekształcenia. Na szczególną uwagę

zasługuje fakt, iż metody, które opierają się atakom pojedynczym wykazują się zmniejszoną odpornością na atak kombinowany

5. Podsumowanie

Istnieje obecnie wiele metod znakowania wodnego, które mają służyć ochronie własności intelektualnej. Wydaje się jednak, że nie istnieje jedna metoda, która odznaczałaby się uniwersalnym przeznaczeniem oraz jednakowo dobrą odpornością na najpopularniejsze ataki. W niniejszym artykule nie przedstawiono rozmyślnych działań mających na celu usunięcie bądź zatarcie znaku wodnego, natomiast skupiono się przede wszystkim na atakach wykorzystujących powszechne przekształcenia o z góry dobranych parametrach. Wyniki mogą pozostawiać jednak zbyt wiele do życzenia w kwestii odporności znaków wodnych i ich praktycznego, skutecznego wykorzystania, otwierając tym samym pole do dalszych badań.

Bibliografia

[1] Bruyndonckx O., Quisquater J., Benoit M. *Spatial method for copyright labeling of digital images.* IEEE Workshop on Nonlinear Signal and Image Processing. CityplaceThessaloniki, s. 456-459. 1995.

[2] Chodkowski A. *Encyklopedia muzyki.* Wydawnictwo Naukowe PWN, 2006.

[3] Chun-Shien Lu (red.) *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property*, Idea Group Publishing, 2004.

[4] Cox placeI. J., Miller M. L., Bloom J. A. *Digital Watermarking.* Morgan Kaufman Publishers, San Francisco 2002.

[5] Kim J. R., Moon Y. S. *A robust wavelet-based digital watermarking using level-adaptive thresholding*, Proceedings of the International Conference on Image Processing, 1999, Vol. 2, s 226-230.

[6] Piva A.Barni M., Bartolini F., Cappellini V. *DCT-based watermark recovering without restoring to the uncorrupted original image.* International Conference on Image Processing, vol. 111, pp. 520-523, 1997.

[7] Rosa L. *DCT-based watermark recovering without resorting to the uncorrupted original image* [online]. *http://www.advancedsourcecode.com/dctwater.asp* [dostęp: 2007].