



Wyższa Szkoła Technologii
Informatycznych w Katowicach

REFERAT PRACY DYPLOMOWEJ

Temat pracy: Analiza wybranych metod ukrywania danych w obrazach cyfrowych

Autor: Krzysztof Kubicius

Promotor: dr hab. inż. Mariusz Boryczka

Kategorie: informatyka, ochrona informacji

Słowa kluczowe: steganografia, ukrywanie danych

1. Cel i podstawowe założenia

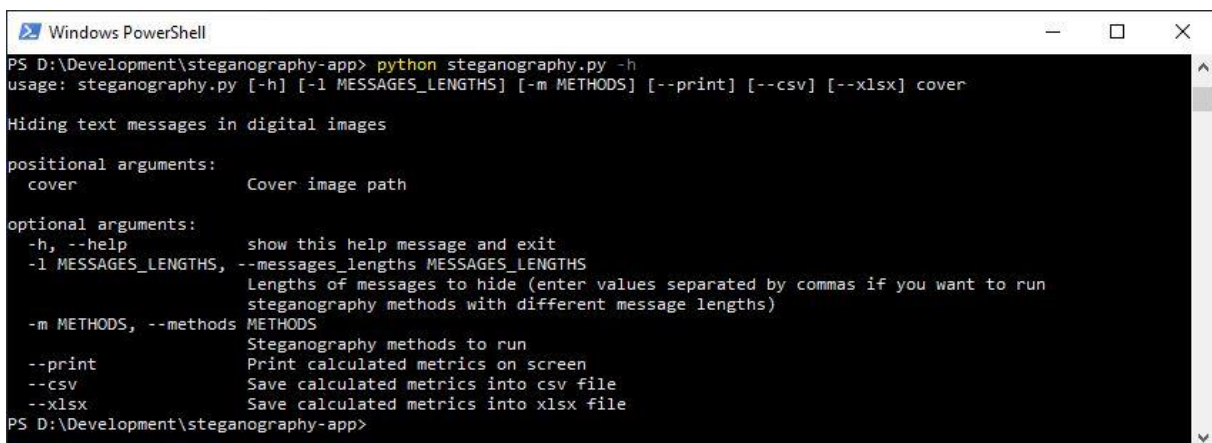
Celem pracy była analiza wybranych metod służących do ukrywania danych w obrazach cyfrowych. Stosowanie takich nośników informacji pozwala na ukrywanie dużej ilości danych jednocześnie nie wzbudzając podejrzeń o samą fakcję osadzenia wiadomości. Kierując się zamiarem poznania specyfiki i efektywności odmiennych podejść do steganografii obrazów cyfrowych zbadano metodę najmniej znaczącego bitu reprezentującą grupę metod substytucji oraz metodę wykorzystującą szybką transformatę Fouriera, która jest metodą transformacyjną.

2. Realizacja projektu

W pierwszej kolejności w pracy opisano genezę, zastosowania, założenia i wyzwania dotyczące tematyki steganografii. Opisano również najczęściej występujące w literaturze metody steganografii obrazów cyfrowych, a także miary jakości i metryki służące do oceny steganogramów, czyli obrazów z ukrytą wiadomością. Na podstawie tych danych opracowano sposób przeprowadzenia badań oraz dokonano selekcji metod i danych wejściowych.

Podstawą do wysnucia wniosków na temat właściwości i użyteczności wybranych metod było przeprowadzenie dwóch eksperymentów. Jeden z nich miał na celu zbadanie wpływu ukrywania wiadomości tekstowej na jakość obrazów. Drugi służył do zebrania i przeanalizowania danych dotyczących korelacji długości tajnego przekazu i jakości zmodyfikowanych plików graficznych.

W celu przeprowadzenia eksperymentów przygotowano aplikację konsolową umożliwiającą osadzanie tekstu o zadanej długości w obrazach, stosując metodę najmniej znaczącego bitu oraz metodę wykorzystującą szybką transformatę Fouriera. Kolejnym jej zastosowaniem jest obliczanie wyników metryk błędu średniokwadratowego (MSE), stosunku sygnału do szumu (PSNR) oraz współczynnika podobieństwa strukturalnego (SSIM) potrzebnych do oceny miar jakości. Wśród miar jakości wyróżnia się pojemność, odporność na modyfikacje oraz statystyczną i sensoryczną niewykrywalność steganogramu. Dodatkowo opracowano algorytm generujący histogramy i wizualizacje przydatne do wskazania różnic pomiędzy działaniem poszczególnych metod steganograficznych.



```
Windows PowerShell
PS D:\Development\steganography-app> python steganography.py -h
usage: steganography.py [-h] [-l MESSAGES_LENGTHS] [-m METHODS] [--print] [--csv] [--xlsx] cover

Hiding text messages in digital images

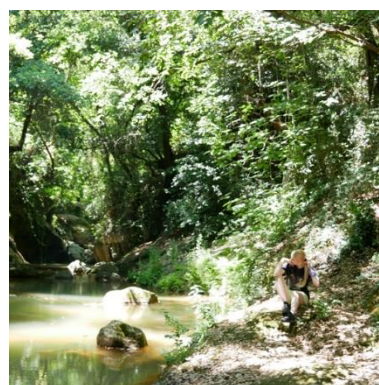
positional arguments:
  cover                Cover image path

optional arguments:
  -h, --help            show this help message and exit
  -l MESSAGES_LENGTHS, --messages_lengths MESSAGES_LENGTHS
                        Lengths of messages to hide (enter values separated by commas if you want to run
                        steganography methods with different message lengths)
  -m METHODS, --methods METHODS
                        Steganography methods to run
  --print               Print calculated metrics on screen
  --csv                 Save calculated metrics into csv file
  --xlsx               Save calculated metrics into xlsx file
PS D:\Development\steganography-app>
```

Skrypt steganography.py wywołany z parametrem -h wyświetlający listę parametrów

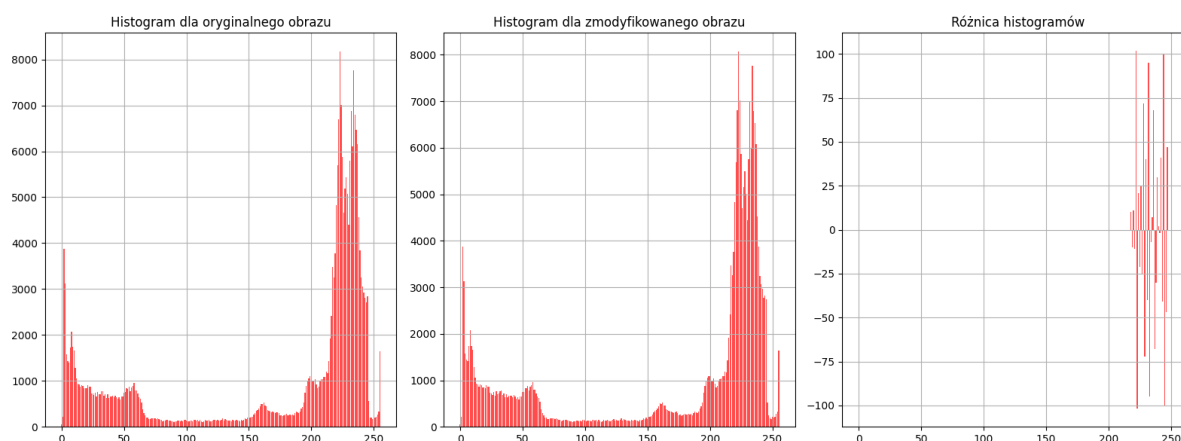
3. Produkt końcowy – wyniki eksperymentów

W ramach eksperymentu dotyczącego wpływu ukrywania wiadomości tekstowej na jakość obrazów w trzech obrazach o rozdzielczości 512 na 512 pikseli ukryto tekst o długości 2048 znaków za pomocą wybranych metod, po czym wyliczono wyniki metryk dla utworzonych steganogramów. Wyniki wykazały, że w przypadku metody najmniej znaczącego bitu (LSB) wpływ na jakość obrazu jest niewielki i nie jest zależny od nośnika. Metoda wykorzystująca transformatę Fouriera (FFT) osiągnęła lepsze wyniki w 2 z 3 obrazów, zatem jest to metoda, której wyniki powiązane są z właściwościami obrazu będącego nośnikiem ukrywanych informacji.



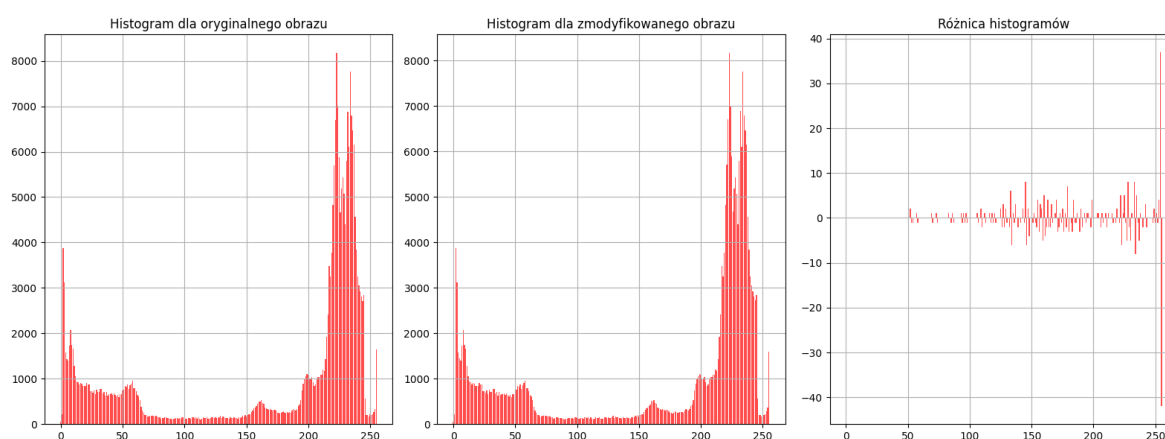
Obrazki „figurka”, „kameleon”, „las” użyte w eksperymentach

W przypadku obrazu przedstawiającego figurkę wykonane histogramy pozwalają zauważyć różnicę jaka istnieje pomiędzy obiema metodami. W przypadku histogramów dla metody LSB widoczne jest, że obraz doznał istotnej zmiany jedynie w ramach najwyższych wartości, podczas gdy dla niższych ani jedna wartość nie zmieniła liczby wystąpień. Powodem jest fakt, że metoda LSB polega na nieznacznej modyfikacji pikseli zaczynając od piksela znajdującego się w lewym górnym rogu obrazu systematycznie wypełniając kolejne rzędy do momentu, gdy zostanie osadzona cała wiadomość. Poniżej zaprezentowano histogramy wyłącznie dla koloru czerwonego, jednak w pozostałych przypadkach histogramy pozwalają na podobną obserwację.



Histogramy kanału czerwonego dla obrazu "figurka" - przed i po osadzeniu wiadomości metodą LSB

W przypadku metody FFT na histogramie widać, że w odróżnieniu od metody LSB, wprowadza modyfikacje w obrębie całego obrazu. Kanały przedstawione na histogramach pokazują, że dla tego samego obrazu „figurka”, piksele zostały zmodyfikowane w zupełnie inny sposób. Modyfikacja fazy obrazu zamiast zmiany wartości pikseli w określonym szyku sprawia, że finalnie piksele zmodyfikowane są w mniej regularny sposób, co wpływa na większą trudność w odkryciu faktu, że dany plik graficzny jest steganogramem.



Histogramy kanału czerwonego dla obrazu "figurka" - przed i po osadzeniu wiadomości metodą FFT

Różnica pomiędzy metodami LSB i FFT jest jeszcze lepiej widoczna na specjalnie przygotowanych wizualizacjach. Kolorem zielonym zaznaczone

zostały te piksele, które zostały zmodyfikowane względem obrazu oryginalnego. W ten sposób zwizualizowane zostało to, co zostało opisane jako wnioski z analizy histogramów.



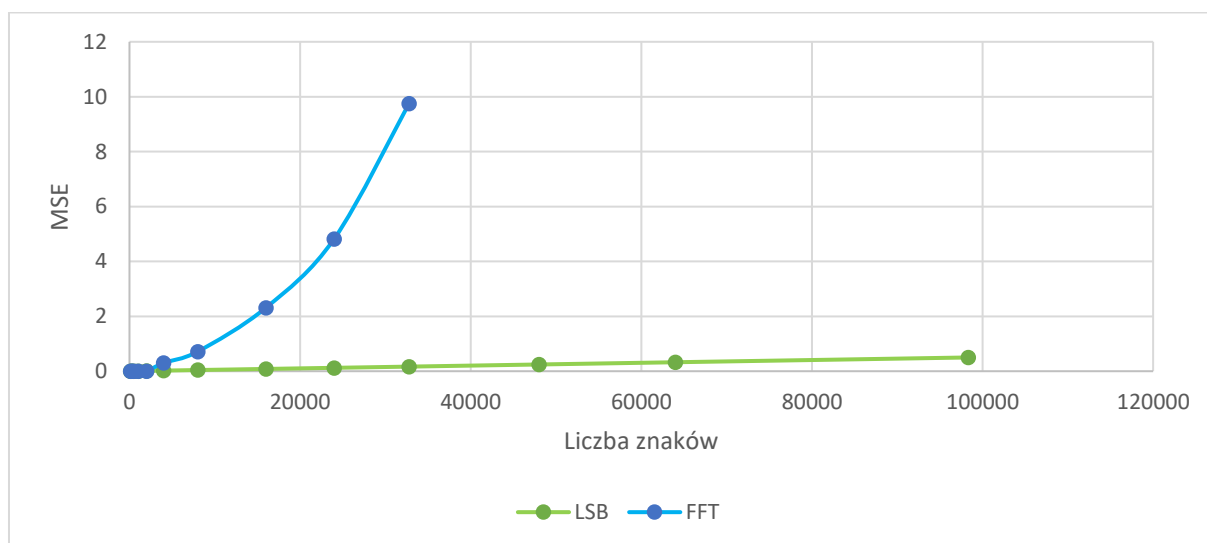
Wizualizacje obszarów zmodyfikowanych przez algorytmy osadzania LSB (po lewej) i FFT (po prawej)

Drugi eksperyment dotyczył korelacji długości osadzonej wiadomości i jakości obrazów. Polegał na utworzeniu steganogramów i zebraniu wyników metryk dla trzynastu różnych długości ukrywanych wiadomości. Dla każdej długości i metody wykonano trzy próby i wyniki uśredniono. Najmniejszą wartością była liczba 125 znaków, natomiast największa wartość różniła się w zależności od maksymalnej pojemności danej metody. Do eksperymentu wybrano obraz „figurka”, którego wyniki metryk w poprzednim eksperymencie dla metody FFT znalazły się w drugim kwartylu, natomiast dla metody LSB wyniki były bardzo zbliżone do pozostałych.

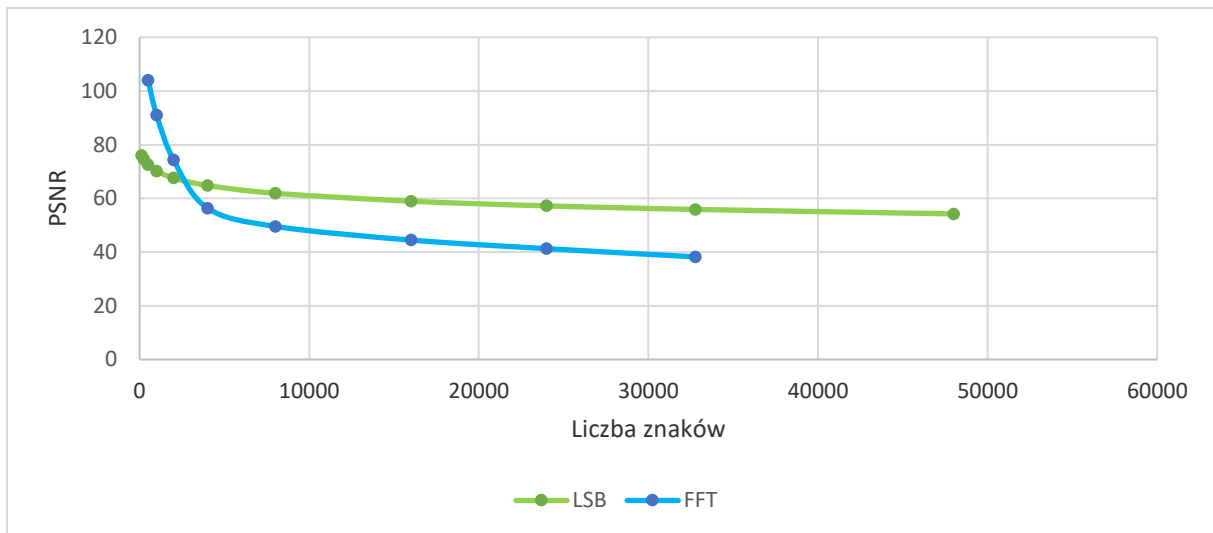
Rezultaty eksperymentu jasno wskazują na bardzo wysoką jakość steganogramów przygotowanych metodą FFT. Pod warunkiem, że liczba znaków wynosi 2000 lub mniej, zmiany wprowadzone przez algorytm osadzania niemalże nie mają wpływu na jakość obrazu. W przypadku dłuższych wiadomości wyniki metryk intensywnie pogarszają się wraz ze wzrostem liczby znaków, na co wskazuje dosyć wysoki błąd średniokwadratowy i niski stosunek sygnału do

szumu, dyskwalifikując metodę FFT jako metodę pierwszego wyboru dla dłuższych tekstów, ze względu na znaczny wpływ na właściwości nośnika. Pomimo tego wyniki nie wskazują na łatwość w dostrzeżeniu zmian podczas analizy sensorycznej.

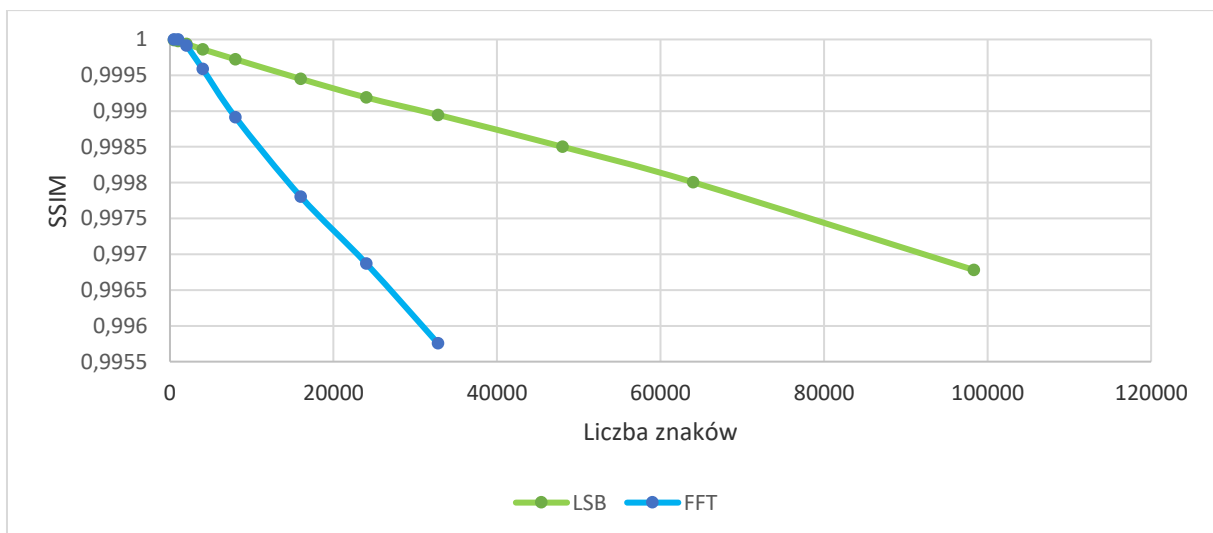
W przypadku metody LSB, mimo gorszych wyników dla krótkich wiadomości, jej zaletą jest zbliżony do liniowego, nieznaczny spadek jakości obrazów wraz ze wzrostem liczby znaków. Skutkuje to tym, że nawet przy maksymalnej długości wiadomości obraz zachowuje relatywnie wysoką jakość. Osiągane w takim przypadku wyniki w dalszym ciągu wskazują na dużą trudność w zauważeniu wprowadzonych modyfikacji przez osobę analizującą steganogram. Wykresy zestawiające wyniki metryk dla obrazów o danej długości zostały zaprezentowane poniżej.



Wykres korelacji wyników metryki MSE w stosunku do długości wiadomości dla metod LSB oraz FFT



Wykres korelacji wyników metryki PSNR w stosunku do długości wiadomości dla metod LSB oraz FFT



Wykres korelacji wyników metryki SSIM w stosunku do długości wiadomości dla metod LSB oraz FFT

Wyniki eksperymentów wykazały, że pod względem wpływu algorytmów osadzania – LSB oraz wykorzystującego FFT – na jakość obrazu nie można jednoznacznie stwierdzić, czy bardziej odpowiednia jest metoda substytucji, czy metoda transformacyjna. Metoda najmniej znaczącego bitu cechuje się większą pojemnością, poza tym daje dobre wyniki niezależnie od treści i nośnika, a także jest prosta w implementacji. Metoda FFT osiąga bardzo dobre wyniki w przypadku krótszych tekstów i odpowiednio dobranego nośnika.

Dodatkowym atutem jest nieregularne rozmieszczenie informacji – co utrudnia wykrycie – oraz duża odporność na modyfikacje pliku.

Wybór metody steganograficznej powinien być dopasowany do indywidualnych potrzeb. W przypadku, gdy dane są wrażliwe i istnieje wysoka szansa, że steganogram zostanie poddany stegoanalizie, a jednocześnie przekaz nie jest zbyt długi, może być to przesłanka, aby zastosować metodę FFT, jednak wiąże się to z dużym nakładem pracy i koniecznością dobrego zrozumienia tematu. Jeżeli ryzyko próby wykrycia steganogramu jest niskie lub ukryte dane nie są wrażliwe, a chodzi głównie o ich przetransportowanie w niewzbudzającej podejrzeń formie, metoda LSB powinna być właściwym rozwiązaniem pod warunkiem, że plik nie zostanie poddany dalszej modyfikacji.

4. Informacje o możliwości wykorzystania / wykorzystaniu pracy

Praca może zostać wykorzystana jako baza wiedzy na temat steganografii obrazów cyfrowych, a zwłaszcza metod steganograficznych badanych w ramach pracy. Może stanowić podstawę do zbudowania szerszego zestawu danych na temat efektywności zróżnicowanych algorytmów osadzania i wyodrębniania informacji z plików graficznych. Badanie takiego obszaru pozwala na opracowywanie rozwiązań zapewniających bezpieczeństwo danych.