

Metody Obliczeniowe w Nauce i Technice

Laboratorium 9

Zastosowania DFT

Patryk Wojtyczek

1 Twierdzenie o splocie

W zadaniach realizowanych na laboratorium korzystamy z jednego fundamentalnego twierdzenia - twierdzenia o splocie, które brzmi: "Transformata splotu dwóch sygnałów (funkcji) jest zwykłym iloczynem ich transformat." Ścisłe rzecz biorąc,

$$\mathcal{F}\{x(t) * y(t)\} = \mathcal{F}\{x(t)\}\mathcal{F}\{y(t)\}$$

\mathcal{F} to oznaczenie dla transformaty Fouriera, a splot definiujemy jako

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) \cdot g(t - \tau) d\tau$$

Dla przypadku dyskretnego

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m] \cdot g[n - m]$$

Korelację z kolei definiujemy bardzo podobnie

$$(f \star g)(t) = \int_{-\infty}^{\infty} f(\tau) \cdot g(t + \tau) d\tau$$

Dla przypadku dyskretnego

$$(f \star g)[n] = \sum_{m=-\infty}^{\infty} f[m] \cdot g[n + m]$$

W zadaniach intersowała nas korelacja (cross-correlation), więc aby zastosować tutaj ww twierdzenie o splocie, trzeba było dokonać spostrzeżenia:

Całka z definicji korelacji jest równoważna całce z definicji splotu, jeśli jeden z sygnałów jest sprzężony (należy pamiętać, że sygnały mogą być zespolone choć nie mieliśmy tutaj z takimi doczynienia) i odwrócony w czasie.

Ścisłe, chcemy więc obliczyć

$$(x \star y)(t) = x(t) * y^*(-t)$$

Sprzężenie w dziedzinie częstotliwości jest równoważne odwróceniu czasu w dziedzinie czasu (z def transformaty Fouriera), więc

$$\mathcal{F}\{x(t) * y^*(-t)\} = \mathcal{F}\{x(t)\}(\mathcal{F}\{y(t)\})^* = \mathcal{F}\{x(t)\}\mathcal{F}\{y(-t)\}$$

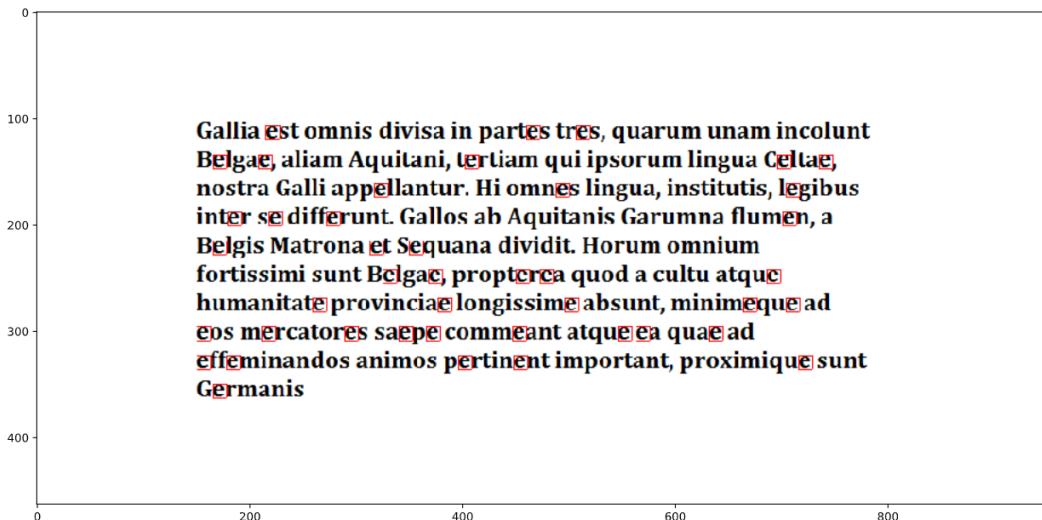
Powyzsza równość tłumaczy naszą magiczną linijkę:

```
1 np.real(ifft2(fft2(img) * fft2(np.rot90(template, 2), img.shape)))
```

A oprócz tego mówi, że alternatywnie do odwracania jednego z sygnałów mogliśmy wziąć sprzężenie już po obliczeniu jego transformaty. Warto też dodać, że dodajemy brakujące zera do sygnału aby nie obliczyć splotu cyklicznego (czy właściwie upewnić się, że obliczony splot kołowy jest równy liniowemu splotowi).

2 Analiza obrazów

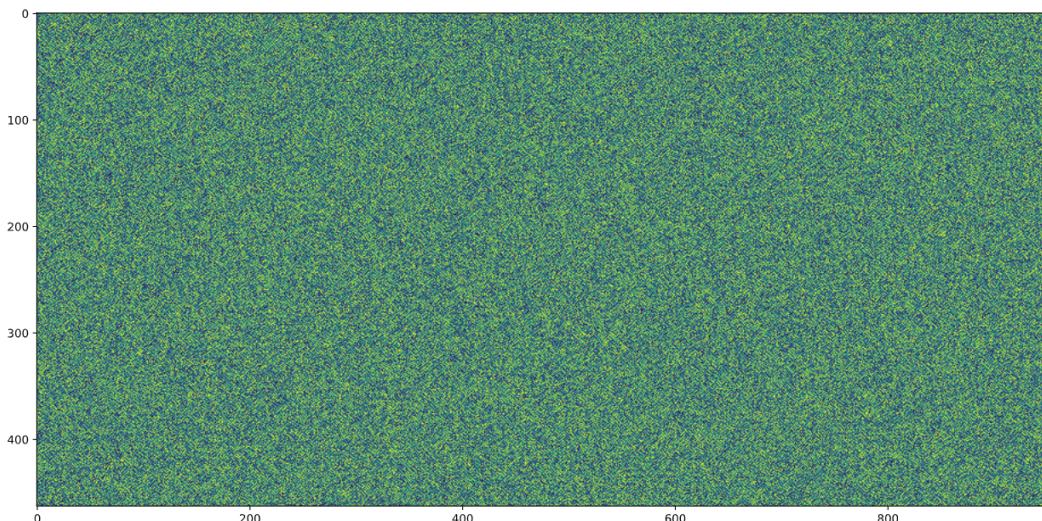
Celem zadnia było znalezienie liczby wystąpień wzorca. Korzystając z wyżej omówionych matematycznych zawiłości, znalezienie liczby wystąpień literki 'e' nie było trudne - wystarczyło wybrać miejsca w których znormalizowana wartość korelacji (dla zdjeć po odwróceniu kolorów i usunięciu szumu) była większa niż pewien próg (przy czym jego dobry wybór nie miał aż tak dużego znaczenia bo wzorzec doskonale pasował)



Rysunek 1: Znalezione wzorce

Możemy przyjrzeć się bliżej znalezionym współczynnikom Fouriera tekstu wyżej

Faza

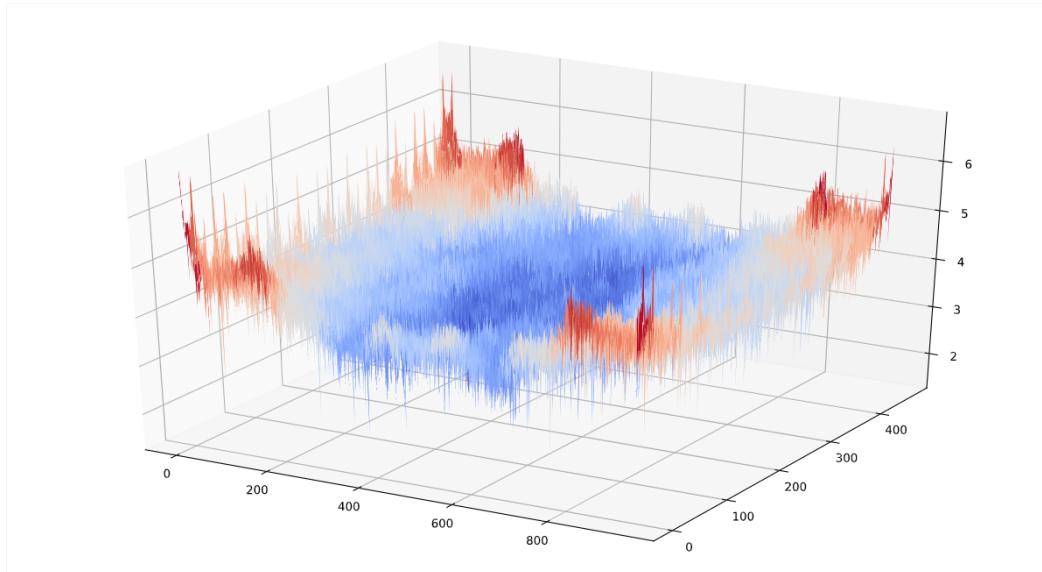


Rysunek 2: Wartości fazy współczynników Fouriera

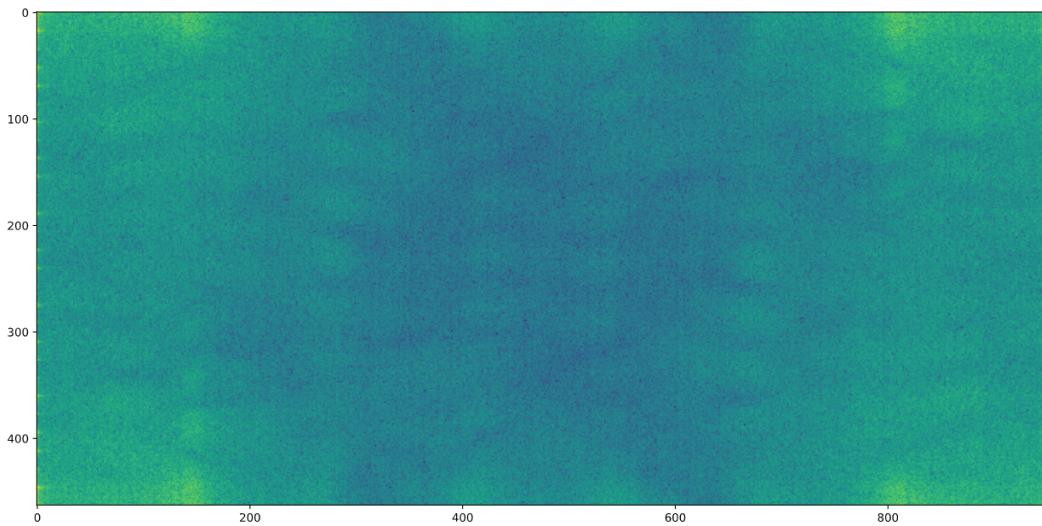
Cóż z tego zdjęcia ciężko wyciągnąć konkretne informacje. Natomiast faza współczynników Fouriera może być (ponoć) wykorzystana do wykrywania krawędzi w aperiodycznym zdjęciu. Warto też dodać, że istnieje

technika wykrywania względnego przesunięcia dwóch zdjęć do której wykorzystuje się korelację faz (phase correlation).

Amplituda



Rysunek 3: Logarytm amplitudy 3D



Rysunek 4: Logarytm amplitudy 2D

Ten rysunek jest dużo bardziej treściwy. Pokazuje jakich sinusoid jest w naszym sygnale (zdjęciu) najwięcej. Widać jasne pionowe prążki (pionowe sinusoidy) sugerujące orientację tekstu. Sugeruje to w oczywisty sposób, że możemy wykorzystać DFT do wykrycia rotacji tekstu.

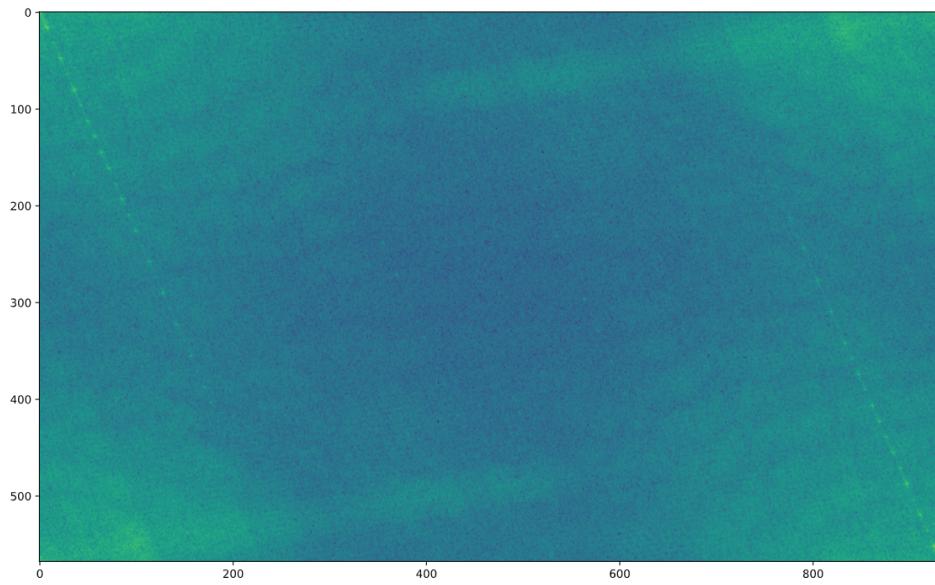
Przekrzywiony tekst

Poniżej zamieszczam przykład bardziej dobrze ilustrujący fakt omówiony wyżej.

Three Rings for the Elven-kings under the sky,
Seven for the Dwarf-lords in their halls of stone,
Nine for Mortal Men, doomed to die,
One for the Dark Lord on his dark throne
In the Land of Mordor where the Shadows lie.
One Ring to rule them all, One Ring to find them,
One Ring to bring them all and in the darkness bind them.
In the Land of Mordor where the Shadows lie.

Rysunek 5: Słynny cytat z Władcy Pierścieni

Amplituda

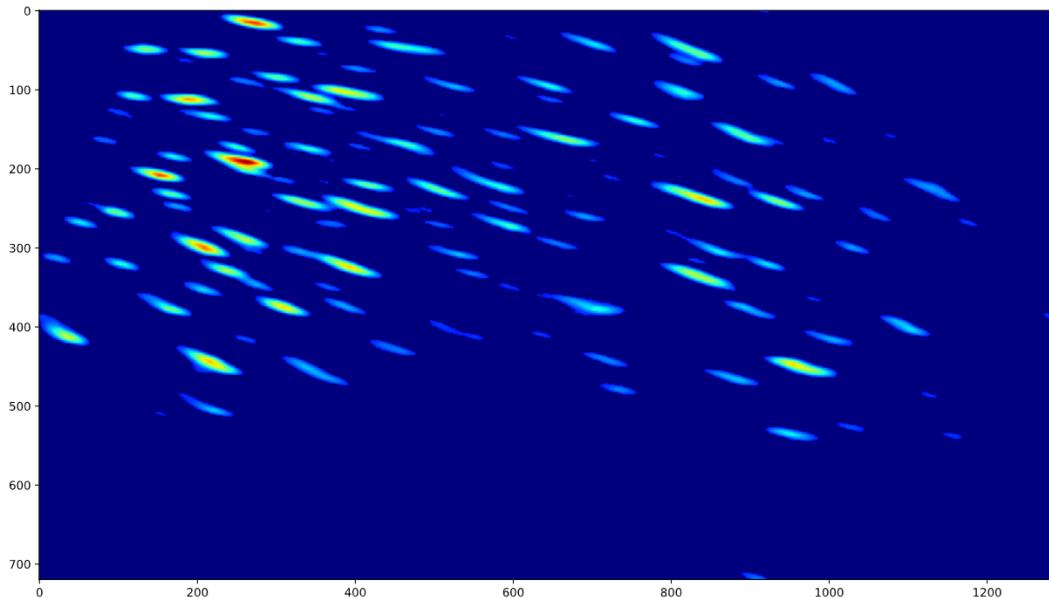


Rysunek 6: Logarytm amplitudy

Widać jasne prążki, które są skierowane pod tym samym kątem co obrócony tekst. Podsumowując, głównie (tzn. ma największą amplitudę) na zdjęciu tekstu składa się periodyczny sygnał(y) o kierunku zgodnym z rotacją tekstu.

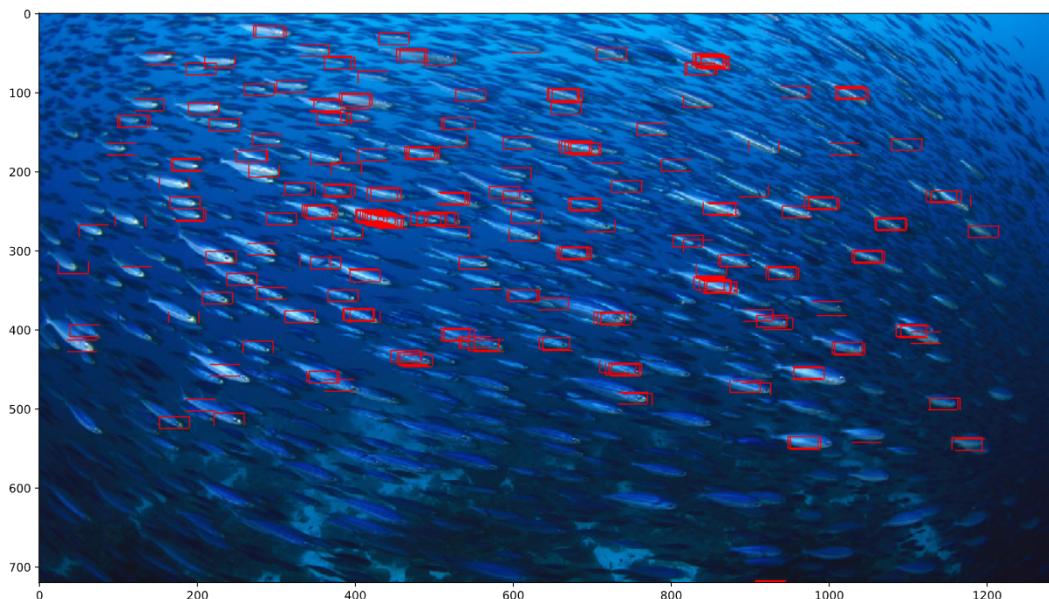
Ławica

Tutaj sprawa była już trudniejsza. Można było zaobserwować, że zdjęcie składa się z wszechobecnego koloru niebieskiego i zielonego oraz w niewielkich ilościach z czerwonego. Rozsądnie zatem było wybrać kolor czerwony gdyż był on czynnikiem różnicującym poszczególne fragmenty zdjęcia.



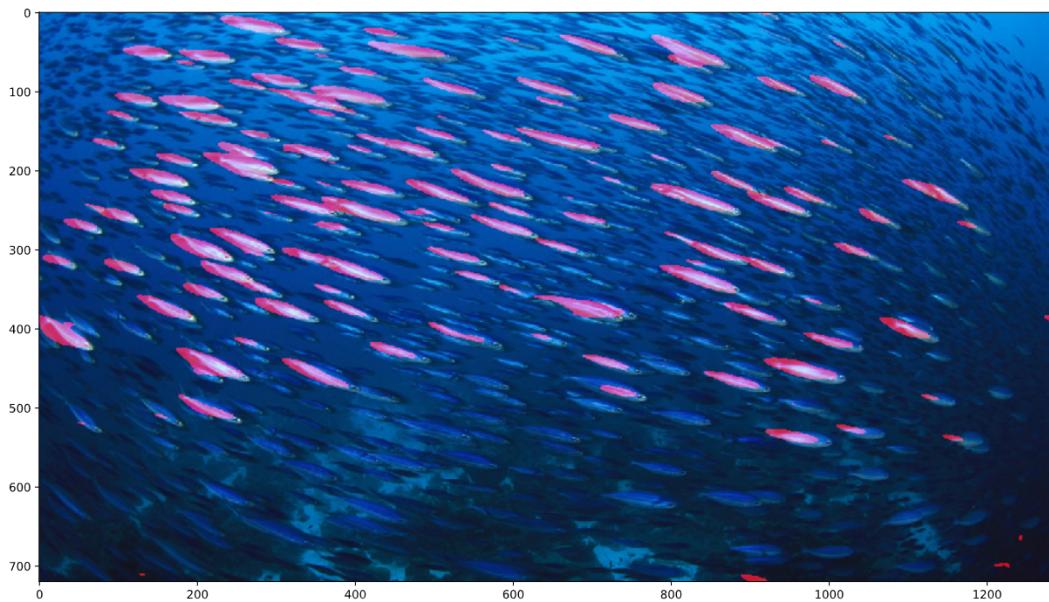
Rysunek 7: Korelacja

Wybierając lokalne maxima lokalne z otrzymanej korelacji, można zlokalizować niektóre wystąpienia wzorca



Rysunek 8: Nałożone wzorce

Sytuacja jest niestety beznadziejna bo wzorzec ma jeden stały rozmiar, a rybki jakkolwiek podobne są różnego rozmiaru. Być może lepsze było takie dobranie progu aby zaznaczać pojedyńcze pixele i w ten sposób uchwycić wzorce (Cóż przynajmniej wizualnie).



Rysunek 9: Nałożone pojedyńcze pixele

Dwa przykłady pokazują zalety i wady stosowania dft do template matchingu. Widać, że gdy jakość jest dobra i istnieją jednoznaczne matche, dft radzi sobie dobrze. Natomiast gdy tak jak w drugim przypadku całość jest bardzo do siebie podobna i wzorce są różnych rozmiarów naiwne szukanie maximów korelacji nie wystarcza. Problemy mogą pojawić się także gdy na naszym zdjęciu nie będzie żadnego wystąpienia template'u - wtedy przy złym doborze progu możemy sklasyfikować niewłaściwie obiekty jako match.

3 Optical Character Recognition

Dyskretna transformata Fouriera jest wszechobecna w technice. Ironicznie raczej rzadko stosuje się ją do tworzenia OCRów, stąd odpowiednie dobranie parametrów było dość wymagającym zadaniem.

Samo wybieranie maximów lokalnych korelacji dla kolejnych sprawdzanych znaków działa słabo (bardzo) z kilku powodów:

- wszechobecne kropki
- gdy w tekście nie było literki, której szukamy mogliśmy zakwalifikować literkę, która była inną literką (i być może już wcześniej ją zakwalifikowaliśmy) i nadpisywaliśmy rezulat. Dzieje się tak dlatego, że próg bazujący na maksymalnej znalezionej korelacji tak naprawdę przepuszcza też miejsca które pasują słabo ale względnie do reszty dobrze, a takich jest dużo. W ten sposób w wyżej pokazanym cytacie z Władcy Pierścieni miałem 37 wystąpień 'Q', gdzie w rzeczywistości nie było tam żadnego.
- analizując znaki można też było zauważyc, że często jest tak, że jedna literka 'zawiera' inną literkę. To oczywiście zależy od czcionki ale często tak jest dla 'e' i 'c', 'T' i 'I', 'm' i 2 razy 'n'. To też jest problem bo mogliśmy w ten sposób nadpisać poprzednio poprawnie zakwalifikowane literki (np zastąpić faktyczne wystąpienie 'e' przez 'c')
- w jakiś sposób musimy uzupełnić znaki białe których oczywiście nie znajdziemy używając korelacji.

Jak żyć?

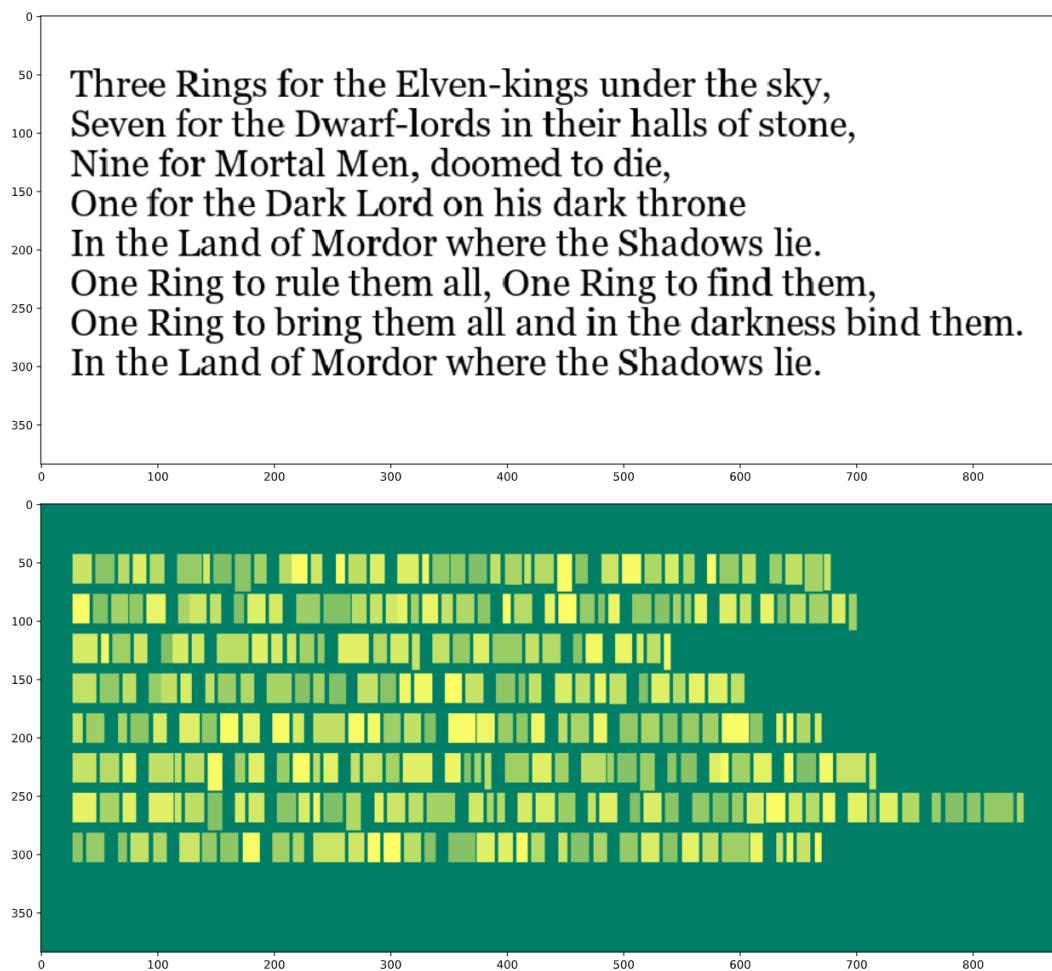
Otoż aby rozwiązać problem nadpisywania już zmatchowanych znaków moglibyśmy trzymać maskę o rozmiarach zdjęcia, która mówiła czy dana pozycja jest już zajęta. Samo w sobie nie rozwiązywało to wszystkich problemów bo zamykamy się wtedy na złe matche o których pisałem wyżej - czyli znajdując literkę c, która pasuje dobrze (a nawet bardzo jak się okazuje ale jest złym matchem) na miejsce literki e, a potem próbując zmatchować literkę która pasuje doskonnale - 'e' nie będziemy mogli tego zrobić.

Teraz znowu; jak żyć?

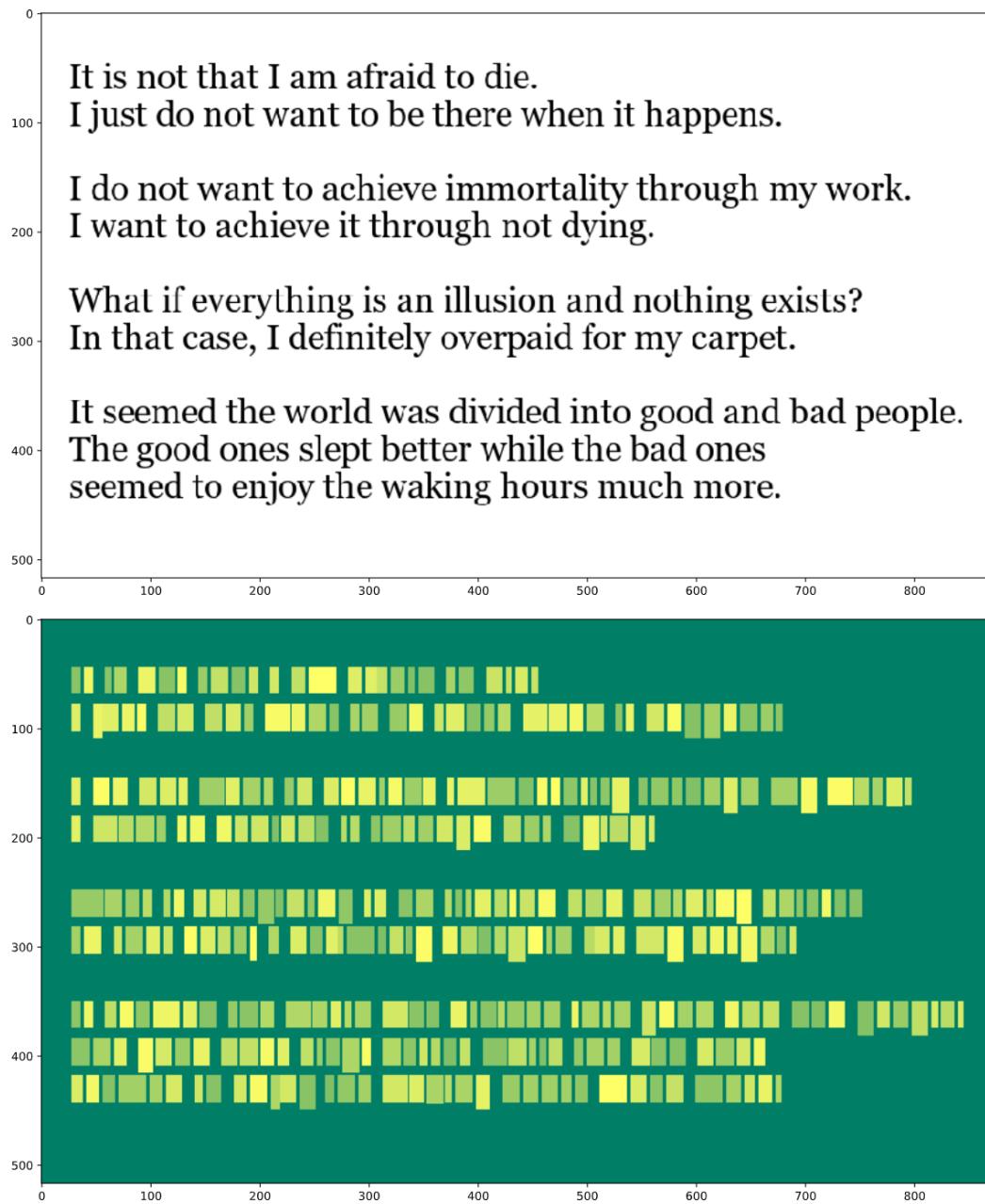
Możemy skorzystać z obserwacji opisanej wyżej - niektóre templatey zawierają inne wewnętrz siebie. Przeglądając się bliżej możemy dojść do wniosku, że tylko templatey które mają więcej białych pixeli mogą zawierać inne wewnętrz siebie. W takim razie ustawiając kolejność w jakiej będziemy przykładać templatey do zdjęcia możemy zadbać o to aby najpierw brać te które potencjalnie mogą zawierać inne wewnętrz siebie - to w dużej mierze eliminuje problem mismatchu, a nawet kropek które składają się z niewielu (chyba jednego) białych pixeli, bo będą one przyporządkowywane na szarym końcu w momencie gdy pozycje z poprawnymi znakami będą już zajęte

Z tak dobranymi heurystykami jedyne co pozostało to odpowiednie dobranie wartości thresholdów i napisanie czegoś co odtworzy białe znaki. Cóż z thresholdem nie ma rady trzeba eksperymentować aby dojść do czegoś co pasuje dla różnych czcionek, natomiast białe znaki można w bardzo prosty sposób odtworzyć ze względu na dokładność wybierania maximów z korelacji (pokazaną poniżej). Wystarczy na podstawie rozmiaru czcionki szacować szerokość spacji i pomiędzy kolejnymi znakami w przypadku odpowiednio dużej przerwy je umieszczać (być może wiele). Podobnie można postąpić ze znakami nowej linii.

Przykłady



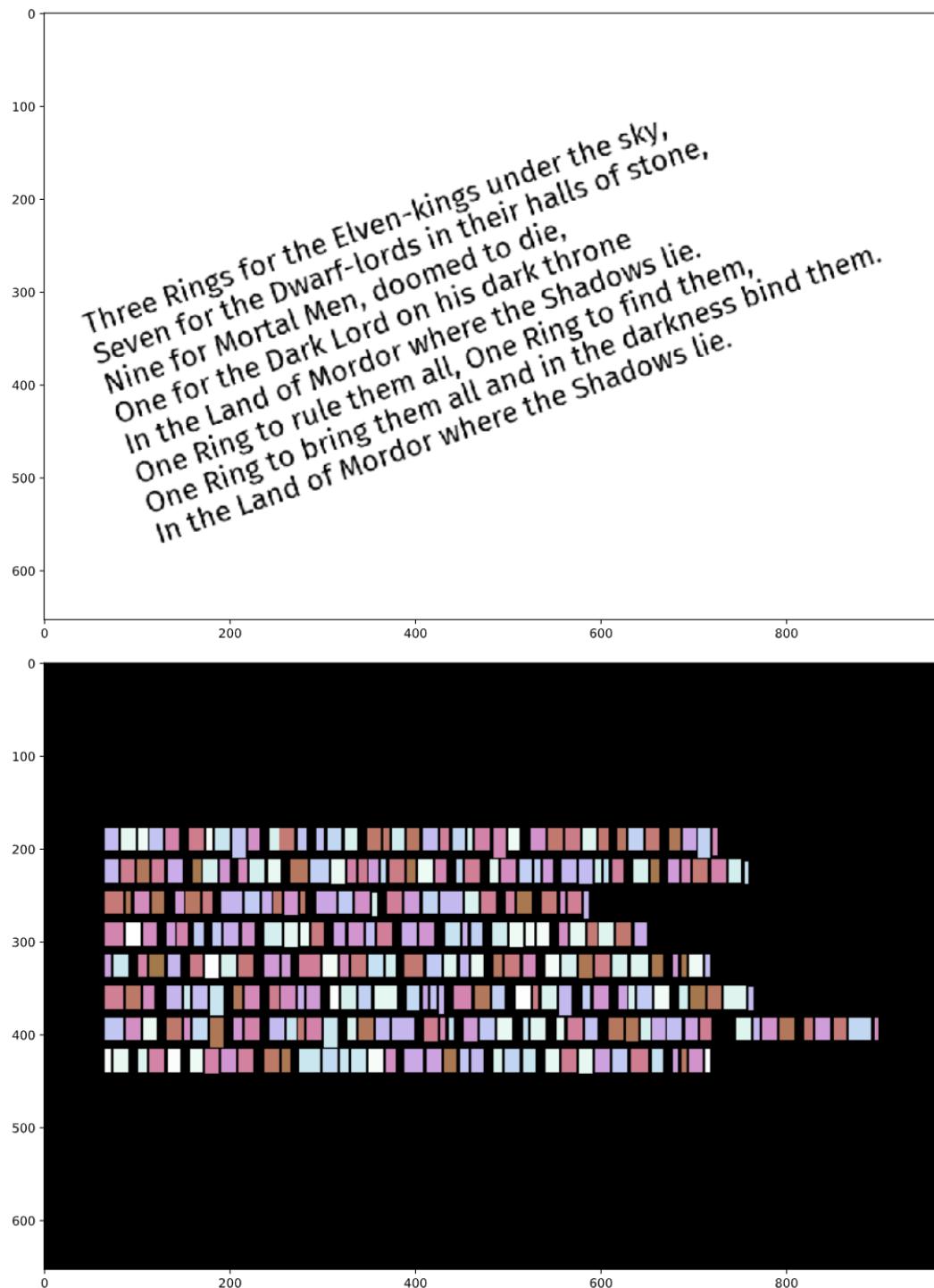
Rysunek 10: Gegorgia, bez rotacji



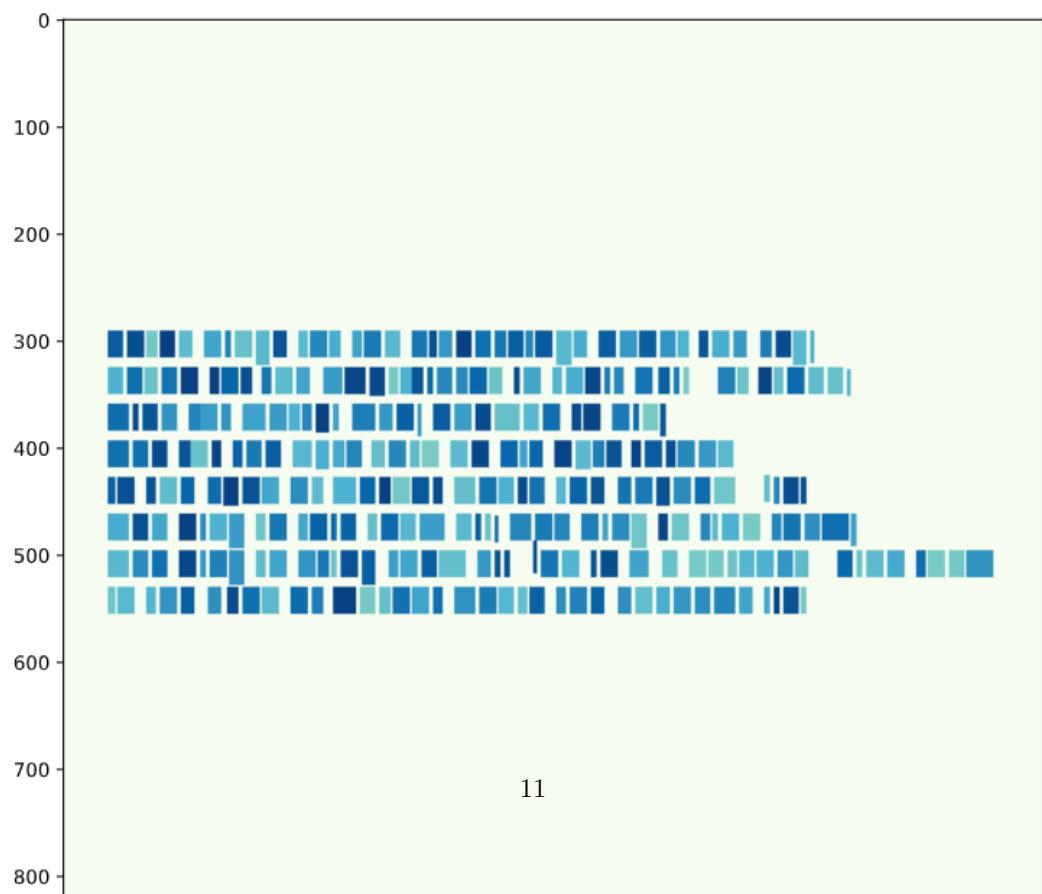
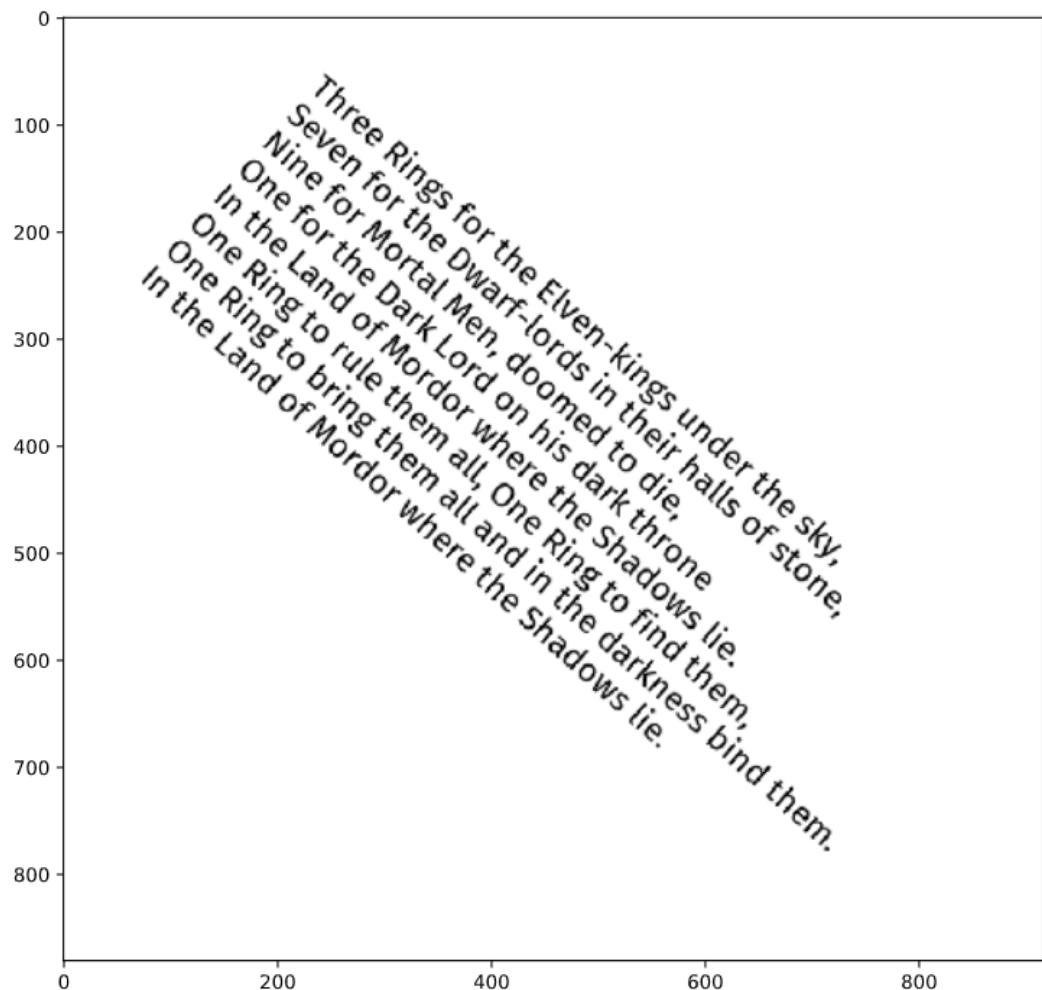
Rysunek 11: Gegorgia, bez rotacji

Pokazane tutaj dwa teksty zostały wyrenderowane z czcionką Georgia (serif), poniżej tekstu widać maskę o której pisałem wcześniej (czyli miejsca do których ocr przyporządkował znaki). Dla obu tych tekstów dokładność wynosi 100%. Oczywiście nie zawsze udaje się osiągnąć taką dokładność, czasem jedne czcionki radzą sobie lepiej od innych. Natomiast z pewnością dla tekstu, który jest dobrego jakości (czyli nie został poddany rotacjom etc) metoda sprawdza się dobrze (wręcz zaskakująco dobrze jak na prostotę działania).

Bardziej problematyczne przykłady



Rysunek 12: Firasans, kąt -30°



Powyżej pokazałem bardziej problematyczne przykłady, widać że jakość nie jest już tak dobra co będzie powodować problemy przy dopasowywaniu templateów.

Jako, że dokładność nie jest 100%-towa zamieszczę znaleziony tekst:

Pierwszy tekst, firasans kąt -30°, dokładność 97%.

Three Rings for the Elven-kings under the sky
Seven for the Dwarf-lords in their halls of stone,
Nine for Mortal Men, doomed to die,
One for the Dark Lord on his dark throne
In the Land of Mordor where the Shadows lie
One Ring to rule them all, One Ring to find them
One Ring to bring them all and in the darkness bind them
In the Land of Mordor where the Shadows lie

Drugi tekst, firasans kąt 42°, dokładność 85%.

Three Rings for the Elven-kings under the sky,
Seven for the Dwarf-lords in their hall of stone
Nine for Mortal Men, doomed to die,
One for the Dark Lord on his dark throne
I
One Ring to rule them all One Ring to find them,
One Ring to bring them all and in the darkness bind them
In the Land of Mordor where the Shadows lie.

Wniosek jest taki, że wraz z pogarszaniem się jakości zdjęcia efektywność OCRA spada ze względu na to, że templatey coraz mniej pasują do tego co jest na zdjęciu.

W załączonym kodzie źródłowym przygotowałem notebook (zarówno do pierwszego jak i drugiego zadania) gdzie znajduje się więcej (znacznie) więcej przykładów działania pod różnymi kątami i dla wszystkich trzech czcionek:

1. Georgia (serif)
2. Firasans (sans-serif)
3. Verdana (sans-serif)

Założyłem jeden rozmiar czcionki (zdefiniowany gdzieś w kodzie źródłowym) choć zmiana nie powinna mieć wpływu na skuteczność (jedynie na rozmiar renderowanych zdjęć). Dokumentację programu również zamieściłem przy kodzie źródłowym więc nie będę jej tutaj przytaczał.

Podsumowując dla zdjęć o dobrej jakości ocr działa zaskakująco dobrze, natomiast im mniejsza jakość zdjęcia tym gorszych rezultatów należy się spodziewać.

Podsumowanie

Dyskretna transformata Fouriera w połączeniu z algorytmami do szybkiego jej obliczania to jedne z najbardziej fundamentalnych algorytmów (myślę, że to jeden z najważniejszych algorytmów XX wieku), który ma olbrzymią ilość zastosowań, między innymi:

- szybkie obliczanie splotu i korelacji co widzieliśmy na przykładach wyżej (co samo w sobie ma wiele zastosowań)
- podstawowe narzędzie w analizie sygnałów
- szybkie mnożenie wielomianów (splot)
- a także kompresja danych - pliki jpg są kompresowane używając fft.