

Algorytm do oceny przydatności skanów preparatów histopatologicznych

inż. Kamil Kasjan

inż. Wojciech Sopot

Abstrakt—W artykule przedstawiono autorski algorytm służący do oceny jakości skanów preparatów histopatologicznych. W procesie pozyskiwania skanów za pomocą automatycznego mikroskopu, na różnych etapach powstawania obrazu cyfrowego mogą pojawić się nieprawidłowości wynikające zarówno z przyczyn mechanicznych jak i błędów w procesie przesyłu i wstępnej obróbki danych cyfrowych. Opisany algorytm składający się z sekwencji testów pozwala określić czy wykonany skan preparatu jest prawidłowy, czy też należy wykonać go ponownie.

■ **WSTĘP** Rozwój automatyki i informatyki pozwala usprawnić i zautomatyzować czynności w wielu dziedzinach. Również w medycynie znajdują zastosowania nowoczesne urządzenia pozwalające skrócić czas i zwiększać skuteczność diagnostyki. Do takich zadań należy również badanie preparatów histopatologicznych. Obecnie istnieją specjalistyczne mikroskopy, które w sposób automatyczny wykonują m.in. skany preparatów umożliwiając patomorfologowi obserwację tkanek w dużym powiększeniu na ekranie komputera bez konieczności ręcznej obsługi mikroskopu. Jednym z problemów, które pojawiają się w procesie digitalizacji obrazów preparatów jest budowanie obrazu o bardzo dużej rozdzielczości poprzez łączenie zdjęć małych fragmentów danego preparatu i późniejszym ich łączeniu. Wymaga to mechanicznej synchronizacji całego układu optycznego oraz połączenia tzw. kafelków w odpowiedniej kolejności. Proces ten nie zawsze kończy się sukcesem i należy każdorazowo

sprawdzać jakość zbudowanego skanu. Do tego celu zaproponowano prosty algorytm bazujący wyłącznie na podstawowych operacjach przetwarzania obrazów w celu uzyskania rozwiązania szybkiego i efektywnego.

■ **OPIS ALGORYTMU** Algorytm można podzielić na kroki, w których na skanie wyznaczany jest obszar zainteresowania tj. obszar gdzie znajduje się tkanka, a następnie przeprowadzane są testy sprawdzające skan pod kątem występowania nieprawidłowości określonego typu. Na podstawie obserwacji kilkuset testowych skanów ustalono, że należą do nich:

- braki fragmentów obrazu
- nieprawidłowo wykonane sklepanie kafelków w skanie
- przeświecenie lub niedoświecenie obrazu

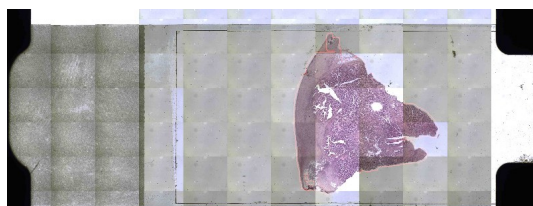
Gdy badany skan zostanie określony jako prawidłowy, obszar na którym znajduje się tkanka zostaje wyodrębniony i możliwe jest przystąpienie do obserwacji.

Wyznaczenie regionu zainteresowania

Przed wykonaniem serii testów, konieczne jest określenie regionu zainteresowania (ROI), czyli obszaru w którym na obrazie znajdują się tkanki. Występowanie błędów w skanie nie musi go dyskwalifikować pod warunkiem, że nie znajdują się one w ROI. Ponieważ tkanki w preparatach są barwione, łatwo można określić ROI na podstawie analizy koloru. W algorytmie dokonuje się transformacji obrazu z przestrzeni RGB do przestrzeni HSV. Następnie wykonywana jest operacja podwójnego progowania. Pikselom o kolorach z zakresu $[126, 180]$ przypisywana jest wartość 255, a pozostałym wartość 0. Aby zlikwidować pojedyncze piksele, którym przypisano wartości 255 na obrazie przeprowadza się operację erozji, a następnie dylatacji w celu przywrócenia odpowiedniej grubości konturów. Tak przygotowany obraz jest poddawany działaniu filtra splotowego, w którym wartości macierzy jądra są następujące:

$$K = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (1)$$

Obraz będący wynikiem ostatniej operacji zawiera kontury wszystkich obszarów skanu, na których znajdują się tkanki.

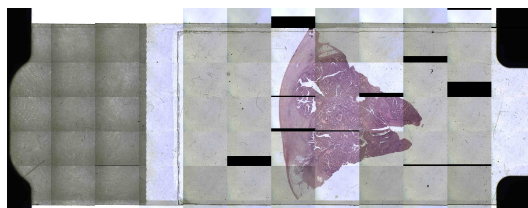


Rysunek 1. Obraz z zaznaczonymi kolorem czerwonym konturami wykrytego obszaru zainteresowania

Wykrywanie brakujących fragmentów obrazu - test 1

Brakujące fragmenty obrazu zawsze objawiają się jako czarne prostokąty (wartość każdego piksela jest równa $RGB(0,0,0)$).

Wykrycie tego typu nieprawidłowości polega na sprawdzaniu wartości kolejnych pikseli, które znajdują się w ROI. Jeśli wartości n sąsiednich pikseli w płaszczyźnie poziomej odpowiadają kolorowi czarnemu, oraz jednolita wyżej opisana sekwencja pikseli pojawia się w kolejnych m wierszach, algorytm ocenia obraz jako uszkodzony i



Rysunek 2. Przykładowy skan powstały z niekompletnych danych

test kończy się negatywnie. Istnieje możliwość, że w obrazie naturalnie pojawiają się obszary całkowicie czarne, dlatego wartości n i m należy dobrać na tyle duże, żeby nie otrzymać wyniku fałszywie pozytywnego.

Ocena poprawności sklejenia fragmentów skanu - test 2

Skan, w którym kafelki nie zostały sklejone w prawidłowej kolejności lub wystąpiły inne okoliczności wpływające na integralność obrazu ma jedną cechę, na podstawie której można ocenić poprawność obrazu. Gdy kafelki nie są odpowiednio sklejone, na obrazie pojawiają się duże, skokowe różnice wartości między pikselami wzdłuż linii pionowych lub poziomych.



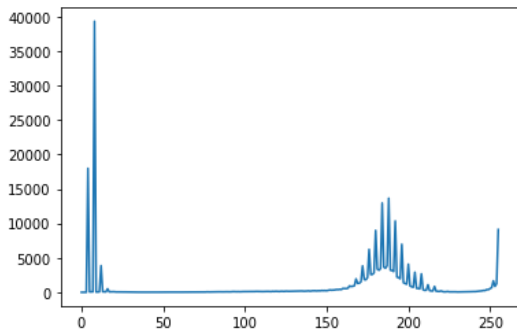
Rysunek 3. Przykład niepoprawnie sklejonego i prześwietlonego skanu

Aby wykryć tego typu wzór, algorytm począwszy od lewego, górnego rogu obrazu sprawdza różnicę wartości między sąsiednimi pikselami w płaszczyźnie horyzontalnej. Jeśli wystąpi różnica większa od ustalonego progu c_{thres} , sprawdzane są pary pikseli znajdujące się w n kolejnych wierszach. Jeśli taka sama skokowa zmiana wartości wystąpi w N_{max} kolejnych wierszach, wykrywana jest nieciągłość wynikająca z nieprawidłowego sklejenia obrazu. Istotny jest dobór prawidłowej wartości parametru N_{max} , ponieważ takie same skokowe zmiany wartości sąsiednich pikseli będą występowały na krawędziach tkanek, a w specyficznych przypadkach krawędzie mogą

być fragmentami równoległe lub prostopadłe do krawędzi obrazu, a wtedy wynik testu mógłby być fałszywie pozytywny.

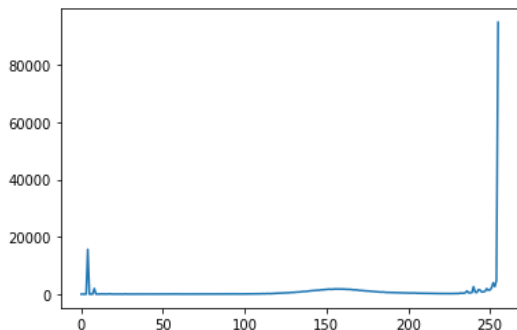
Ocena luminacji obrazu - test 3

Prześwietlenie lub niedoświetlenie skanu można określić na podstawie zbadania histogramu obrazu skonwertowanego do skali odcieni szarości. Występowanie nadmiernej liczby wartości mniejszych od dolnego progu jasności oznacza że obraz jest niedoświetlony.



Rysunek 4. Histogram skanu niedoświetlonego

Analogicznie pojawianie się zbyt dużej liczby wartości większych od górnego progu jasności pozwala na stwierdzenie że obraz jest prześwietlony.



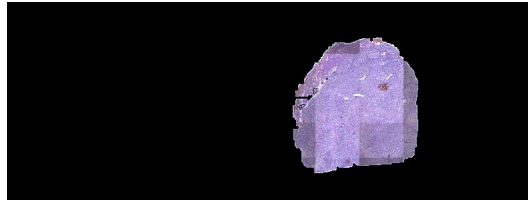
Rysunek 5. Histogram skanu prześwietlonego

Przyjęto, że jeśli w obrazie piksele o wartościach ze skrajnych przedziałów wyznaczonych przez progi przekracza 5% to obraz może być źle oświetlony i test daje wynik negatywny.

Ekstrakcja ROI z obrazów

Jeśli wszystkie testy się powiedą, w algorytmie usuwane są zbędne sektory nie wnoszące

żadnych informacji. Odbywa się to poprzez wypełnienie wszystkich obszarów poza ROI kolorem czarnym. Taki obraz jest wynikiem działania algorytmu i może być przekazany dalej jako prawidłowy.



Rysunek 6. Prawidłowy obraz będący wynikiem działania algorytmu

IMPLEMENTACJA I DZIAŁANIE ALGORYTMU

Narzędzia programistyczne

Program do oceny skanów napisano w języku Python oraz wykorzystano bibliotekę OpenCV 3. Zastosowanie tego języka pozwoliło na minimalizację ilości kodu potrzebnego do wykonania zadania. Biblioteka OpenCV dostarcza gotowych funkcji umożliwiających przeprowadzanie operacji na obrazach takich jak:

- filtracja splotowa
- erozja i dylatacja
- transformacja przestrzeni barw
- składanie kilku obrazów w jeden
- wypełnianie określonej przestrzeni na obrazie

Ponadto użyto biblioteki EasyGUI dostarczającej funkcji do budowania graficznych interfejsów użytkownika. Za jej pomocą utworzono prosty eksplorator plików pozwalający wybrać skan do przeprowadzenia testów.

Działanie programu

Program z punktu widzenia użytkownika jest bardzo prosty. Po uruchomieniu pojawia się eksplorator plików, za pomocą którego należy wybrać odpowiedni obraz. Następnie przeprowadzana jest seria testów. Informacje o wyniku testu są przedstawiane jako komunikaty tekstowe w konsoli. Jeśli choć jeden test da wynik negatywny, skan uznawany jest za uszkodzony, o czym użytkownik zostaje poinformowany stosownym komunikatem. W przypadku gdy skan przejdzie pomyślnie wszystkie testy, generowany jest odpowiedni komunikat oraz wyświetlany jest obraz, na

którym znajdują się wycięte fragmenty zawierające tkanki z preparatów.

■ **WNIOSKI I PODSUMOWANIE** Opisany algorytm pozwala w krótkim czasie automatycznie skontrolować poprawność pozyskiwanych danych wizualnych bez użycia skomplikowanych narzędzi oraz wydajnych zasobów sprzętowych. Zastosowanie takiego algorytmu jest dość wąskie, ponieważ w obecnej formie można go wykorzystać wyłącznie w urządzeniach mikroskopowych, jednak dzięki prostocie wykonywanych operacji mógłby być w prosty sposób zaadaptowany do innych zadań.

■ Literatura

1. A. Kaehler, G. R. Bradski, *OpenCV 3 Komputerowe rozpoznawanie obrazu w C++ przy użyciu biblioteki OpenCV*, wydawnictwo Helion, Warszawa, 2018
2. A. Urbaniak, *Metoda detekcji tkanki na preparatach histopatologicznych*, praca magisterska, Wydział Informatyki i Zarządzania, Politechnika Wrocławska, 2018