JAKOŚĆ OBRYSÓW EKSPERCKICH W OCENIE SEGMENTACJI

QUALITY OF EXPERTS CONTURS IN VERIFICATION OF SEGMENTATION

Jan Juszczyk, Silesian University of Technology
(maj 2011, prof. dr hab. inż. Ewa Piętka)

Abstract

The article describes a method of how some experts determine the counters of structures on radiological images. The counters determined by experts are used as a point of reference for automatic and semiautomatic methods of segmentation. That's why the study of experts' counters quality is also presented in this article. The quality study was carried out on a group of 6 people doing some works connected with drawing. This group was chosen because of the greatest experience in drawing and so very likely in marking structures counters. The study was done to generally investigate the range of mistake generated by a human. The results revealed that experts' mistake is comparable or in some cases even bigger than the mistake generated by automatic or semiautomatic methods presented in professional literature. It is shown in the article how to initially estimate the value of experts' counters as a point of reference for other methods.

Streszczenie

Artykuł opisuje sposób wyznaczania eksperckich obrysów struktur na obrazach radiologicznych. W artykule przedstawiony jest również sposób badania jakości obrysów eksperckich używanych jako metoda referencyjna dla półautomatycznych i automatycznych metod segmentacji. Badania jakości obrysów zostały wykonane na grupie 6 osób zajmujących się pracami związanymi z rysunkiem. Taka grupa zawodowa została wybrana ze względu na największe doświadczenie związane z wykonywaniem rysunków, co za tym idzie z założenia powinni najdokładniej je wykonywać. Przedstawione badania są badaniami pilotażowymi mającymi za zadanie zbadać ogólnie wielkość błędu generowanego przez człowieka niezależnie od interpretacji obrazu. Badania te wykazały, że wartość błędu generowanego przez eksperta jest porównywalna a czasem nawet większa niż wartość błędu dla metod półautomatycznych i automatycznych przedstawionych w literaturze. Zaproponowana została metoda wstępnego badania jakości obrysów eksperckich pozwalająca określić ich przydatność jako metody referencyjnej.

1. Wstęp

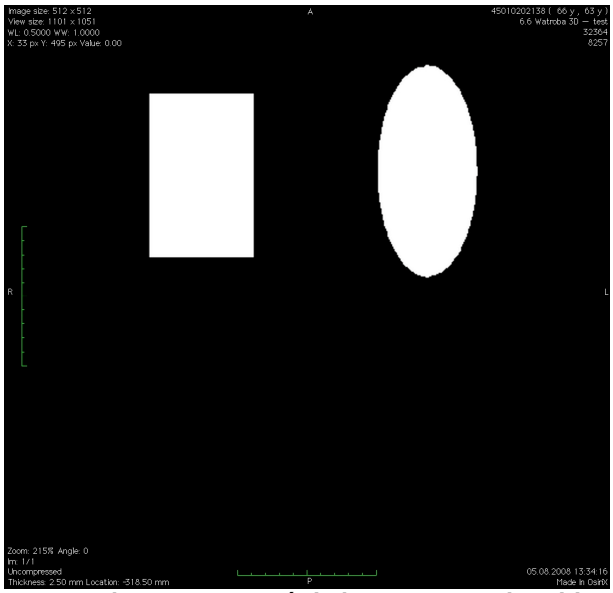
Każda nowa metoda segmentacji, niezależnie od tego do czego ma być użyta, wymaga określenia skuteczności i jaka jest jej niepewność. Każda metoda segmentacji prezentowana w literaturze jest w jakiś sposób oceniania pod względem jakości. Metody umożliwiające przetestowanie metody są oparte na porównaniu uzyskanej objętości do objętości referencyjnej [1-3], porównaniu krawędzi uzyskanej struktury w stosunku do krawędzi zaznaczonej przez eksperta [4] lub ocenie za pomocą skali określającej słownie jakość lub przydatność do dalszej analizy wyniku segmentacji [5,6].

Łatwo zauważyć, że większość z tych metod opiera się na określeniu różnicy między wynikiem segmentacji a danymi referencyjnymi. W przypadku testowania algorytmu na modelu można dokładnie określić jak bardzo wynik segmentacji różni się od oryginału i takie określenie jakości jest jak najbardziej wiarygodne. Problem pojawia się, gdy eksperci wykonujący obrysy nie są ze sobą zgodni. Przyczyn takiej sytuacji można poszukiwać w dwóch czynnikach. Pierwszym z nich jest różna interpretacja przez ekspertów obrazu radiologicznego. Drugim czynnikiem jest niedoskonałość obrysu wynikająca ze zdolności manualnych eksperta. Czynnik ten jest niezależny od kompetencji eksperta jako osoby interpretującej obraz. Aby rozdzielić te dwa czynniki należało wybrać taki obraz, którego interpretacja jest oczywista.

2. Przebieg eksperymentu

W eksperymencie wykorzystano obraz syntetyczny przedstawiający biały prostokąt oraz elipsę na czarnym tle (Rys 1.). Obraz miał charakter binarny. Piksele o wartości 0 reprezentowały tło, a o wartości 1 obiekt. Dzięki temu krawędzie figur geometrycznych były idealnie ostre i nie występowa-

ło żadne rozmycie w rastrze obrazu. Obraz zapisany został w formacie DICOM. Rozmiar rastra wynosił 512x512 pikseli, co jest najpopularniejszym rozmiarem dla cyfrowych obrazów radiologicznych. Obraz wyświetlany był ekspertom w programie OsiriX z wykorzystaniem pełnego dostępnego zakresu jasności.



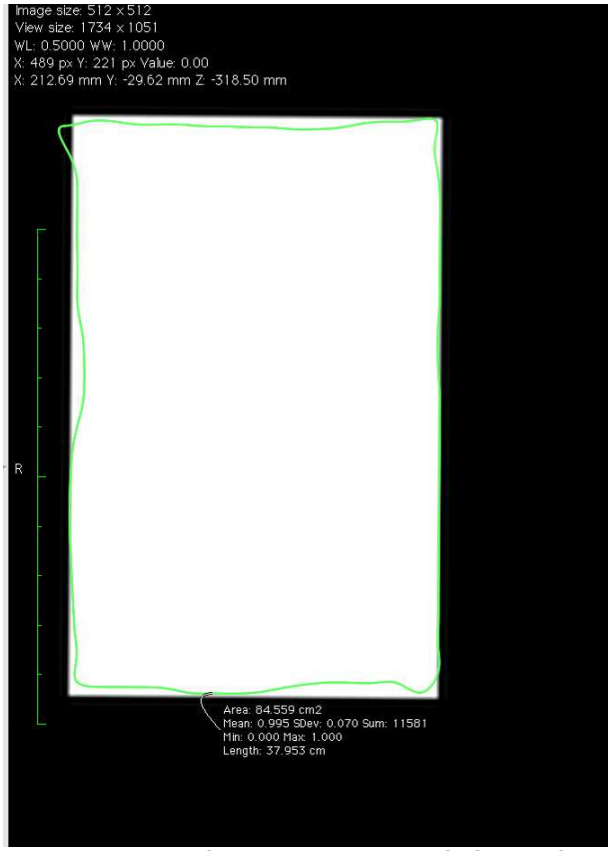
Rys. 1. Obraz testowy wyświetlony w programie OsiriX
Fig. 1. The test image displayed in OsiriX

W eksperymencie wzięło udział sześciu ekspertów zawodowo związanych z odręcznym wykonywaniem rysunków. Każdy z ekspertów miał za zadanie pięciokrotnie wykonać obrys dwóch figur geometrycznych bez możliwości dokonania poprawy wykonanego obrysu (Rys. 2) oraz jeden z możliwością nanoszenia dowolnej liczby poprawek. Eksperci mieli nieograniczony czas na wykonanie obrysów. Obrysy wykonywane były przy użyciu wbudowanych narzędzi programu OsiriX za pomocą tabletu z piórkiem. Ostatni obrys – z możliwością poprawy – najczęściej korygowany był przy użyciu narzędzia *Repulsor* pozwalającego na płynne przesuwanie i dopasowywanie już istniejącego obrysu. Obraz syntetyczny wyświetlany był zawsze w tej samej wielkości dla każdego eksperta, bez możliwości zmiany skali obrazu. Celem eksperymentu było określenie granicznych wartości precyzji wykonania obrysu przez eksperta przy wykluczeniu błędu związanego z jego interpretacją.

3. Wyniki

W wyniku eksperymentu otrzymano trzydzieści sześć obrysów obiektów testowych; dwadzieścia pięć obrysów bez poprawek oraz sześć obrysów dodatkowo poprawianych przez ekspertów. Uzyskane obrysy eksperckie były następnie przetwarzane na obraz DICOM. Obrazy te poddawane były analizie.

Proces analizy uzyskanych obrazów był identyczny jak proces analizy wyników segmentacji dwuwymiarowej. Narzędzia wykorzystane do oszacowania błędu łatwo dają się rozszerzyć na przypadki obrazów trójwymiarowych.



Rys. 2. Fragment obrazu testowego z naniesionym obrysem eksperckim.
Fig. 2. Part of the test image with experts manual contour

W pierwszej kolejności określono względną różnicę objętości DV. Zdefiniowaną w [6]:
„Niech W_{wyn} i W_{odn} będą zbiorami, których elementy to trójki uporządkowane określające położenie w przestrzeni punktów należących odpowiednio do wysegmentowanego obiektu (wynikowego) i obiektu odniesienia. Niech W_{rozn} będzie zbiorem zawierającym elementy, które nie należą do iloczynu zbiorów W_{wyn} i W_{odn}

$$W_{rozn} = \left\{ w : w \in \left(W_{wyn} \div W_{odn} \right) \right\}, \qquad (1)$$

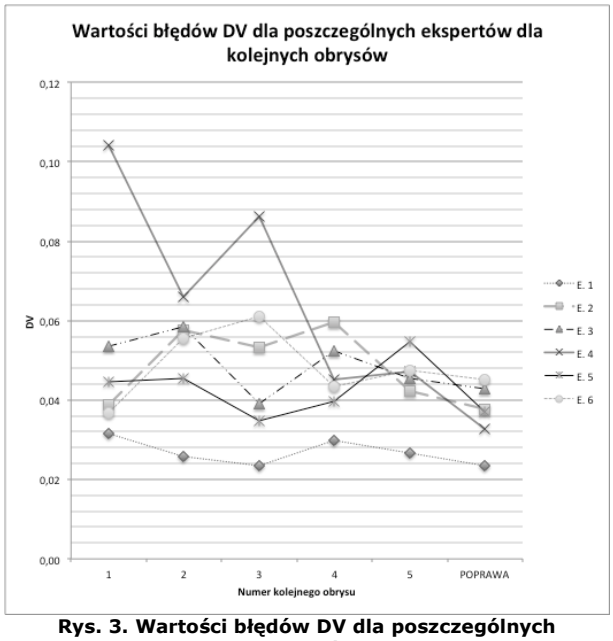
wtedy względna różnica objętości DV dana jest wzorem

$$DV = \frac{\overline{\overline{W_{rozn}}}}{\overline{\overline{W_{odn}}}}, \qquad (2)$$

gdzie $\overline{\overline{W_{rozn}}}$ i $\overline{\overline{W_{odn}}}$ oznaczają moc zbioru odpowiednio W_{rozn} i W_{odn}.”

Parametr DV uwzględnia wszystkie różnice między obrazem oryginalnym a wynikiem eksperckim. W tym konkretnym przypadku jest to lepsze rozwiązanie niż liczenie SVO [1] czy wyznaczanie samej różnicy objętości [7], ponieważ obraz odniesienia jest syntetycznym modelem i jest dokładnie znany więc nie ma sensu odnosić się do obszaru „uśrednionego". W celu określenia granicznej wartości jakości obrysów eksperckich wyznaczono średnią wartość DV ze wszystkich uzyskanych DV – SDV, za wyjątkiem pierwszego obrysu dla każdego eksperta. Wartość ta wyniosła SDV = 0,045 ± 0,014.

Decyzja o odrzuceniu pierwszego wykonanego przez eksperta obrysu została podjęta ze względu na wstąpienie w niektórych przypadkach dużej wartości DV, spowodowanej prawdopodobnie nieprzyzwyczajaniem się eksperta do czułości narzędzia. (Rys. 3.)



Rys. 3. Wartości błędów DV dla poszczególnych ekspertów.
Fig. 3. DV (Difference Volume) for experts and next contours

Proces przystosowywania się eksperta do narzędzia może wymagać wykonania większej liczby obrysów. Mimo wahań wartości DV dla kolejnych obrysów, dla wartości średniej (uśrednianie po ekspercie) DV obserwowana jest tendencja spadkowa. (Rys. 4.). Co ciekawe poprawa wykonanego obrysu nie zawsze dawała lepsze rezultaty. Jeszcze bardziej zaskakujące jest to, że w 50% przypadków najlepszym obrysem jest obrys wykonany jako trzeci.

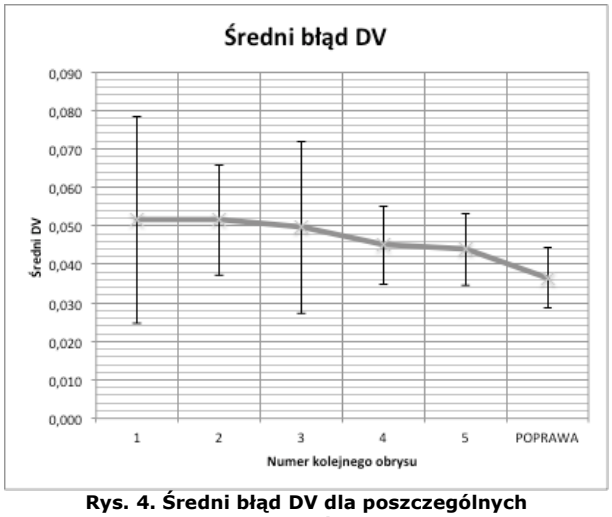
Oprócz wyznaczenia DV każdego z ekspertów zbadano również czułość i specyficzność. Czułość została zdefiniowana jako

$$SEN = \frac{\left|V_{org} \cap V_{\exp}\right|}{\left|V_{org}\right|}, \qquad (3)$$

a specyficzność

$$SPEC = \frac{\left|V_t\right|}{\left|V_t \text{È} \left(V_{\exp} \setminus V_{org}\right)\right|}, \qquad (4)$$

gdzie V_{org} oznacza obiekt (jako całość - prostokąt i elipsa) oryginalny; V_{exp} oznacza obiekt uzyskany przez eksperta, a V_t oznacza obszar tła.

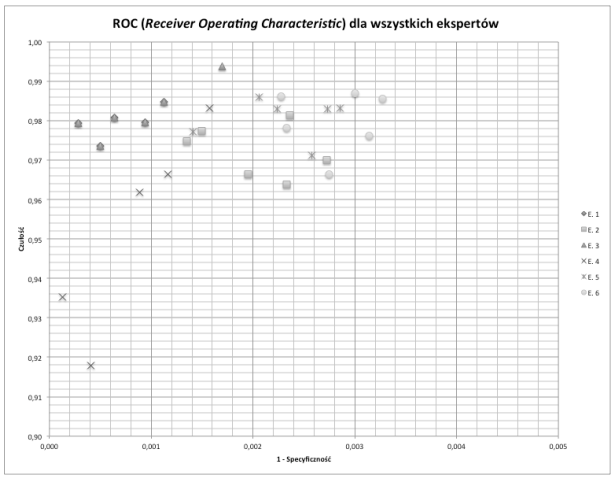


Rys. 4. Średni błąd DV dla poszczególnych obrysów
Fig. 4. Means of DV for manuals

Eksperci uzyskali bardzo wysokie wartości i czułości i specyficzności średnio na poziomie SEN = 0,975 ± 0,016 dla czułości i SPEC = 0,998 ± 0,001 dla specyficzności. (Rys. 5.) W przypadku definicji specyficzności opierającej się na obszarze tła jako wartości prawdziwie pozytywnej ($TP = V_t$), istotne znaczenie ma stosunek wielkości tła do wielkości obiektu segmentowanego. Przy bardzo dużym obszarze tła wyniki mogą być zafałszowane. Należy zatem ograniczyć obszar tła do takiej wielkości jaka faktycznie jest brana pod uwagę w procesie analizy (segmentacji) obrazu. W przypadku wykonywania obrysów przez ekspertów obszarem poddawanym analizie jest cały obszar, ponieważ ekspert został tylko poinformowany o tym, że ma zaznaczyć obiekty na obrazie. Wybór co jest obiektem a co tłem zależał tylko i wyłącznie od eksperta.

4. Weryfikacja eksperta

Proponuje się, aby ekspert wykonujący obrysy referencyjne przed wykonywaniem tych obrysów był przetestowany. Test powinien obejmować wykonanie obrysów na obrazach o pełnej kontrastowości (np. takich jak przedstawione) w celu określenia manualnych zdolności eksperta. W celu poprawy wiarygodności specyficzności obiekty powinny być zróżnicowane pod względem wielkości i losowo rozsiane w polu widzenia.



Rys. 5. Wykres ROC dla wszystkich ekspertów
Fig. 5. ROC for all experts

Ekspert powinien mieć możliwość wcześniejszego wykonania kilku obrysów przed przeprowadzeniem właściwego testu w celu zaznajomienia się z narzędziem do obrysowywania. Proces nauki narzędzia powinien być powtarzany przed każdym wykonywaniem obrysów na właściwych obrazach radiologicznych.

5. Posumowanie

Wyniki przedstawionych badań nie powinny być traktowane jako podstawa do zakwestionowania obrysów eksperckich jako metody referencyjnej dla metod segmentacji. Z punktu widzenia wiarygodności medycznej jedyną osobą, która może określić co jest, a co nie jest poszukiwaną strukturą anatomiczną, jest lekarz radiolog opisujący badanie. Jego obrys z formalnego punktu widzenia jest wynikiem wiążącym. Wszystkie te elementy przemawiają za tym, że mimo wszystko obrysy eksperckie są najlepszą metodą referencyjną w przypadku medycznych obrazów radiologicznych. Uzyskane wyniki należy raczej traktować jako graniczne wartości oceny jakości segmentacji. Oznacza to, że na podstawie obrysów eksperckich nie można powiedzieć że metoda generuje błąd mniejszy niż np. DV=0,05 skoro ekspert uzyskał np. 0,06.

Literatura

1. Campadelli P., Casiraghi E., Pratissoli S., *A segmentation framework for abdominal organs from CT scans* , Artificial Intelligence in Medicine (50) Elsevier, 2010
2. Kayaalp C., Arda K., Oto A., Oran M., *Liver volume measurement by spiral CT An in vitro studyw*, Journal of Clinical Imaging (26) Elsevier, (2002)
3. Lee J., Kim N., Lee H., Seo J. B., Won H. J., Shin Y. M., Shin Y. G., Kim S.-H., *Efficient liver segmentation using a level-set method with optimal detection oft he initial liver boundary from level-set speed images*, Computer Methods and Programs in Biomedicine (88) Elsevier, 2007
4. Campadelli P., Casiraghi E., Pratissoli S., *Liver segmentation from computed tomography scans: A survey and new algorithm* , Artificial Intelligence in Medicine (45) Elsevier, 2009
5. Ruskó L. Bekes G., Fidrich M., *Automatic segmentation oft he liver from multi- and single-phase contrast-enhanced CT images,* Medical Image Analysis (13) Elsevier, 2009
6. Juszczyk J.M., *Półautomatyczna trójwymiarowa segmentacja wątroby i śledziony na podstawie obrazów tomograficznych,* XVII Krajowa Konferencja Biocybernetyka i Inżynieria Biomedyczna, 2011 (w druku)
7. Lim S.-J., Jeong Y.-Y., Ho Y.-S., *Automatic liver segmentation for volume measurement in CT Images,* Journal of Visual Communication & Image Representation Elsevier, 2006

Adres służbowy autora

mgr Jan Juszczyk
Silesian University
of Technology
ul. Akademicka 16
44-100 Gliwice
tel. (32) 237 17 35

jan.juszczyk@polsl.pl