dr hab. inż. Adam Wojciechowski, prof. PŁ
Politechnika Łódzka
Wydział Fizyki Technicznej, informatyki i Matematyki Stosowanej
Instytut Informatyki
90-924 Łódź, ul. Wólczańska 215

Łódź, 29 grudnia 2019

2020 -01- 07
DZIAŁ NAUKI
tel./fax 91449-41-39
te 6-21, 41-96

# Recenzja rozprawy doktorskiej sporządzona dla Senatu Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie

Tytuł rozprawy: **Percepcyjne metody oceny jakości obrazów syntezowanych przez silniki gier komputerowych**

Autor rozprawy: **mgr inż. Rafał Piórkowski**

Promotor: **dr hab. inż. Radosław Mantiuk, prof. ZUT**

***1. Jakie zagadnienie naukowe jest rozpatrywane w prace (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?***

Tematem prac naukowych Autora jest percepcyjna ocena jakości komputerowych obrazów graficznych, syntezowanych z pomocą popularnych silników gier komputerowych. Problem percepcyjnej oceny obrazów graficznych towarzyszy zazwyczaj procesowi konfiguracji ustawień potoku syntezy obrazów. Celem jest bowiem zapewnienie oczekiwanej wydajności procesu syntezy obrazów (zadana liczba klatek na sekundę) przy zachowaniu percepcyjnie najwyższej dostępnej jakości syntezowanej grafiki i przy ograniczonych zasobach sprzętowych (np. gorsza karta graficzna). Zapewnienie odpowiedniej wydajności procesu renderingu obrazów okupione jest zazwyczaj obniżeniem jakości grafiki, w tym w szczególności pogorszeniem: jakości tektur, jakości cieni (przesunięcie, zniekształcenie, itp.), dokładności bufora głębokości, czy jakości anty-aliasingu. Oceny jakości obrazów pozwala na zobiektywizowaną i percepcyjnie Pareto optymalną konfigurację procesu syntezy, zmierzającą do zachowania możliwie najlepszych ustawień jakości grafiki przy równoczesnym wzroście wydajności procesu syntezy obrazów. Wykonanie wspomnianej oceny automatycznie, wymaga jednak zbudowania efektywnych i percepcyjnie wiarygodnych metod oceny jakości obrazu, czemu poświęcił badania Autor rozprawy. Autor nie ograniczył się w swoich badaniach do weryfikacji użyteczności i optymalizacji istniejących miar i metryk oceny jakości obrazów oraz wykrywania powszechnie spotykanych w syntezowanych obrazach artefaktów, ale zaproponował własną metrykę skonstruowaną w oparciu o sztuczną splotową sieć neuronową.

Autor postawił w pracy tezę, że *„obiektywne metody oceny jakości obrazów, rozumiane jako algorytmy komputerowe, są w stanie wykrywać wybrane artefakty i obniżoną jakość na obrazach generowanych przez silniki gier komputerowych oraz oceniać jakość obrazów ze skutecznością zbliżoną do oceny wykonanej przez człowieka. Przy czym jakość rozumiana jest jako pogorszenie wyglądu obrazu mająca na celu zwiększenie wydajności renderingu"*. Zarówno teza jak i sama problematyka sformułowana jest jasno i zwięźle. Problematyka badań naukowych jest ważna i aktualna, gdyż kompromisy w obszarze wydajności i jakości syntezowanych obrazów są powszechnym problemem wszędzie tam gdzie mamy do czynienia z próbami wydajnego i fotorealistycznego renderingu obrazów graficznych, w szczególności w obszarze gier komputerowych i środowisk wirtualnych, jak również podczas optymalizacji czasowo-jakościowej obecnej powszechnie w przemyśle filmowym.

Rozprawa ma charakter eksperymentalno-analityczny. Podstawową metodą badawczą są eksperymenty percepcyjne, prowadzone z udziałem użytkowników, którzy oceniali jakość obrazów syntezowanych przez popularne silniki gier komputerowych. Wyniki oceny zostały poddane analizie statystycznej, a następnie stanowiły referencję dla metod badawczych weryfikujących metody automatycznej oceny jakości obrazów.

Zebrane wyniki badań percepcyjnych posłużyły również do przeprowadzenia kalibracji jednej z metryk oceny jakości obrazów oraz uczenia splotowej sieci neuronowej, stanowiącej autorską metodę oceny jakości obrazu. Wartość rozprawy podnosi uzupełnienie jej o ciekawe przypadki użycia dyskutowanych metod oceny jakości, w szczególności zadanie profilowania *benchmarku* poprzez automatyczny dobór parametrów procesu syntezy obrazów graficznych, stymulowane wskaźnikiem jakości obrazu.

***2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle / świadczący o dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonywujący?***

Charakter eksperymentalno-analityczny rozprawy doktorskiej, osnuty wokół głównego problemu oceny jakości obrazów cyfrowych, determinuje podejście Autora do przeglądu literatury. Przyczynowo-skutkowy, logiczny wywód przedstawiający analizę problemu oceny jakości obrazów, ujęty w kontekście modelowania danych referencyjnych (rozdział 2), metryk oceny jakości obrazów (rozdział 3) i zastosowania splotowych sieci neuronowych do porównywania obrazów (rozdział 4), przeplatany jest celnie wybranymi, choć niezbyt licznymi, pozycjami literatury. Znacznie pełniejszy obraz dokonanego przeglądu literatury można uzyskać sięgając do publikacji naukowych, których doktorant był współautorem (5 pozycji). Wśród nich na szczególną uwagę zasługują publikacje na międzynarodowej konferencji *ACM SIGGRAPH Symposium of Applied Perception* i w renomowanym czasopiśmie *ACM Transactions of Applied Perception*.

Ciekawym uzupełnieniem przeglądu literatury mogłaby być analiza zastosowań referencyjnych metryk oceny jakości obrazów w wybranych obszarach przemysłowych. Autor wprawdzie wspomniał o zagadnieniach kompresji obrazów i wskazał na możliwość zastosowania dyskutowanych metryk w przemyśle gier komputerowych, ale do obszarów, które bezsprzecznie korzystają z podobnych rozwiązań należą: bezpieczeństwo informacji (np. znakowanie wodne obrazów), przemysł filmowy (np. redukcja kosztów produkcji filmowych poprzez uproszczenie nieistotnych percepcyjnie efektów specjalnych), czy rozwijające się dynamicznie środowiska wirtualne i mieszane (np. impozycja syntetycznych obiektów cyfrowych do zdjęć rzeczywistych, wykrywanie błędów w stereo-parach, czy postrzeganie jakości imersyjnych, stereoskopowych obrazów cyfrowych).

Reasumując uważam, że przegląd został wykonany w sposób jasny i przekonywujący, odpowiedni do postawionego problemu naukowego. Krytyczna analiza aktualnego stanu wiedzy pozwoliła nakreślić problematykę prowadzonych badań, wyciągnąć konstruktywne wnioski i zaproponować innowacyjne rozwiązania w obszarze metod oceny jakości obrazów. Bardzo ważnym i godnym podkreślenia obszarem wniosków z przeglądu literatury była zaproponowana przez Autora metodologia badań. Zastosowanie uznanych i popartych światową literaturą metod badawczych, do oceny efektywności metryk/metod oceny jakości obrazów, wzmacnia rangę zaproponowanych rozwiązań.

***3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?***

Autor właściwie rozwiązał postawiony problem naukowy i udowodnił postawioną tezę naukową. Przeprowadził badania użyteczności popularnych i autorskich metod oceny jakości obrazów cyfrowych w zadaniu detekcji artefaktów i zadaniu ich oceny przy różnych poziomach jakości procesu syntezy. Przyjęte w badaniach założenia wynikają bezpośrednio z kontekstu zastosowań, czyli syntezy obrazów cyfrowych przez popularne silniki gier komputerowych w kontekście ograniczonych zasobów sprzętowych.

Bardzo istotnym elementem badań było modelowanie danych referencyjnych, które uwzględniały specyfikę mechanizmów postrzegania obrazów przez system wzrokowy człowieka oraz prawdopodobieństwo poprawności map referencyjnych. Dane te były niezbędne do opracowania i weryfikacji skuteczności metod oceny jakości obrazów cyfrowych. W moim odczuciu bazy danych obrazów zostały przygotowane właściwie,

a eksperymenty percepcyjnej oceny różnic pomiędzy obrazami zostały przeprowadzone wiarygodnie, z dbałością o najwyższe standardy naukowe.

Wśród osiągnięć naukowych zaproponowanych przez Autora, stanowiących element rozwiązania postawionego problemu, należy wskazać oryginalną metodę oceny jakości obrazów, zbudowaną na autorskiej architekturze sieci splotowej. Pierwsza niezależna gałąź sieci splotowej przetwarza obrazy będące różnicami pomiędzy podobrazami obrazów testowych i referencyjnych (uśrednione referencyjne mapy różnic pochodzące z eksperymentów percepcyjnych), zaś druga niezależna gałąź sieci splotowej przetwarza odpowiadające im podobrazy obrazów referencyjnych (obrazy o obniżonej jakości wybranych parametrów renderingu). Architektura sieci jest przekonująca, choć większość parametrów została ustalona empirycznie. Uczenie sieci splotowej odbywało się za pomocą oryginalnie przetworzonych i zinterpretowanych obrazów pochodzących z badań percepcyjnych, przeprowadzonych z udziałem użytkowników. Proces uczenia sieci został przeprowadzony zgodnie z zasadami sztuki. Autor umiejętnie zadbał o osłabienie (niwelowanie) zjawiska przeuczania sieci neuronowej i zdiagnozował problem zrównoważenia liczności klas, któremu próbował zaradzić zwiększając wagi dla wybranych klas. Ciekawym uzupełnieniem dyskusji byłaby analiza „frekwencji" pikseli przynależnych do poszczególnych klas i uzasadnienie merytoryczne zmian wartości wag.

Wynikiem badań literaturowych i przeprowadzonych eksperymentów było opracowanie metodologii zmierzającej do oceny skuteczności metod oceny jakości obrazów cyfrowych z pełną referencją. W rzeczonym kontekście, tytuł rozdziału 5, odnoszącego się do szeroko pojętej wydajności, jest nieco zbyt ogólny – wydajność metod może obejmować również złożoność czasową i obliczeniową metod, które zostały zmarginalizowane.

Na polu licznych eksperymentów Autor wykazał przydatność istniejących klasycznych miar jakości obrazu, w wykrywaniu artefaktów oraz przeprowadził autorską kalibrację metryki SSIM podnosząc jej wynikową skuteczność (wynik AUC) o blisko 5 punktów procentowych. Autor wykazał również przewagę zaproponowanej przez siebie metody oceny jakości obrazu, opartej o splotową sieć neuronową, w zadaniu wykrywania obrazów o obniżonej jakości syntezy grafiki, w porównaniu z konkurencyjnymi rozwiązaniami literaturowymi. Wyniki badań z przeprowadzonych eksperymentów prowadzą do ciekawych wniosków dotyczących wyższości miary L (ang. *Likelihood)* nad miarą pola powierzchni pod krzywą ROC (AUC).

Wnioski z przedstawionych badań jednoznacznie dowodzą, że rozwiązania zaproponowane przez Autora dysertacji są lepsze od istniejących w literaturze przedmiotu, metod i metryk oceny jakości obrazów.

***4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?***

Zgodnie z deklaracją Autora dysertacji, zamieszczoną w podsumowaniu pracy, jego wkład własny obejmuje sześć osiągnięć, z których moim zdaniem najważniejszymi są:

- opracowanie autorskiej metryki oceny jakości obrazów, stosującej splotową sieć neuronową oraz metody polepszania jakości danych uczących;
- przeprowadzenie szczegółowej analizy efektywności referencyjnych metryk oceny jakości obrazów (generowanych przez silniki gier komputerowych) w zadaniach wykrywania artefaktów obrazów cyfrowych i porównywaniu obrazów o różnych poziomach jakości grafiki;

Szczególnie osiągnięcie związane z wieloaspektową analizą efektywności/skuteczności metryk oceny jakości obrazów, zostało przeprowadzone z zachowaniem najwyższych standardów naukowych i niezbicie świadczy o dojrzałości naukowej doktoranta.

Warto również wspomnieć, że przygotowane przez Autora bazy obrazów, przedstawiające artefakty graficzne i obrazy o różnym poziomie jakości cieni, tekstur i metod anty-aliasingu, stanowią bezcenny zbiór danych referencyjnych dla innych naukowców zajmujących się pokrewną tematyką. Brakuje bowiem uznanych,

referencyjnych zbiorów danych, do których powszechnie mogliby odnosić się naukowcy zajmujący się rzeczoną tematyką.

***5. Czy autor wykazał umiejętność poprawnego i przekonywującego przedstawienia uzyskanych przez siebie wyników / zwięzłość, jasność, poprawność redakcyjna rozprawy/?***

Praca została napisana i zredagowana w sposób bardzo staranny. Układ pracy jest logiczny i spójny. Celnie dobrane metody badawcze, w sposób jasny i poprawny dokumentują przewagi zaproponowanych rozwiązań nad referencyjnymi rozwiązaniami literaturowymi. Uzyskane wyniki przedstawione zostały w sposób czytelny i obrazowy.

Odnosząc się do kwestii redakcyjnych nie stwierdzam żadnych istotnych uchybień. Jedynymi uwagami, które nie zmniejszają mojej wysokiej oceny pracy są: nieliczne błędy literowe, np.: „uśredniając" (str. 43); najprawdopodobniej błędnie skonstruowane zdanie na str. 21 *„...obraz testowy zawierał występujący artefakt, natomiast obraz testowy był pozbawiony artefaktu."*; prawdopodobnie błędnie użyty indeks parametru $c_4$ zamiast $c_5$ we wzorze 3.9; słaba czytelność opisów kolumn na rys. 6.2 (słaby kontrast czarnego z granatem); zdublowana pozycja literaturowa w bibliografii [13] i [14];

***6. Jakie są słabe strony rozprawy i jej główne wady?***

Praca nie zawiera żadnych istotnych uchybień i wad. Wszystkie zawarte poniżej uwagi mają charakter polemiczny i nie obniżają wysokiej oceny recenzowanej pracy.

Wśród przyjętej terminologii, pomimo wielu zamieszczonych systematyk i definicji, pojawia się moim zdaniem niekiedy niespójność pojęć tj. „wydajność", czy „metoda vs metryka", które wymagałaby dyskusji. Przykładowo w rozdziale 4.1, wśród metod porównywania obrazów, za pomocą splotowych sieci neuronowych, Autor rozróżnia *ocenę jakości* (pojedyncza wartość liczbowa) od *oceny widoczności* (mapa różnic). Jednocześnie w rozdziale 4.2 wprowadza *metrykę oceny jakości obrazów,* która dostarcza wynikowo mapę różnic zamiast wartości liczbowej. Rozdział 5 dysertacji porusza nominalnie kwestie oceny wydajności dyskutowanych metod. Autor dekomponuje w preambule rozdziału pojęcie „wydajności" na kwestie rzetelności/skuteczności metod oraz kwestie złożoności obliczeniowej i czasowej. Zapisy o wydajności mogą implikować u czytelnika oczekiwania odnośnie analizy złożoności czasowej i pamięciowej, podczas gdy takowych w rozdziale 5 brakuje. Ciekawym uzupełnieniem rozprawy byłoby przeprowadzenie dyskusji w zakresie wydajności (czasowej i obliczeniowej) rozważanych metod lub ograniczenie terminologiczne do „skuteczności" analizowanych metod/metryk.

W rozdziale 2.4, w ramach opisu konstruowania referencyjnych map różnic, Autor zadeklarował stworzenie „łącznie 720 pojedynczych referencyjnych map różnic", jednak z opisu trudno jednoznacznie wywnioskować skąd wzięła się ta liczba.

Zauważalne niedomówienia, w porównaniu do opisów „klasycznych" metryk oceny jakości obrazów, towarzyszą opisom dotyczącym autorskiej splotowej sieci neuronowej. Sieć splotowa (rozdział 4.3) cechuje się deterministycznym zestawem parametrów (ustawień), podczas gdy zarówno jej architektura jak i konfiguracja mogłaby znaleźć pełniejsze uzasadnienie w teorii uczenia maszynowego. Przykładowo, dlaczego wektory wejściowe gałęzi proponowanej sieci (rys. 4.2) zostały podzielone na 3 warstwy (224x224**x3**) skoro mapa różnic jest w skali szarości (10 odcieni szarości odpowiadających 10 klasom). Jednocześnie druga gałąź wymaga prawdopodobnie podziału wektora wejściowego na trzy warstwy (224x224**x3**) ze względu na barwę podobrazów, ale jednocześnie nie specyfikuje modelu barw, w którym interpretowany jest wejściowy podobraz referencyjny. Ciekawe byłoby również uzasadnienie spostrzeżenia dotyczącego procesu nauki sieci, iż „najlepsze wyniki osiągane są w momencie kiedy niewiele większy nacisk położony jest na skraje klasy, a mniejszy na środkowe" (str. 51).

Ciekawym uzupełnieniem dyskusji byłaby odpowiedź na pytanie dlaczego wartości współczynnika L (ang. *Likelihood*), w eksperymentach oceniających skuteczność metryk w wykrywaniu obniżonej jakości grafiki,

wykazują istotnie różne wartości (tabela 5.1) dla jakości tektur i jakości cieni, skoro zgodnie z opisem mają one charakter niskoczęstotliwościowy (str. 62 „odpowiadają jakości cieni i jakości tekstur w niniejszej pracy").

Moim zdaniem bardziej szczegółowych wyjaśnień można by oczekiwać od opisów dotyczących obliczania wskaźników jakości, a szczególnie testowego wskaźnika jakości (rozdział 6.2.2). Chociaż Autor drobiazgowo deklaruje sposób jego wyliczenia na podstawie średniej arytmetycznej z wartości maksymalnej pochodzącej z *testowej mapy różnic* oraz średniej obliczonej ze wszystkich pikseli, to pozostaje pytanie jakie jest uzasadnienie, że otrzymane w ten sposób bezwzględne wartości, będą porównywalne z referencyjnym wskaźnikiem jakości, który został zdefiniowany w dość umowny sposób. Na czym dokładnie polegało zastosowanie błędu średniokwadratowego (ang. MSE) podczas porównania wskaźników jakości (referencyjnego i testowego), które są wartościami liczbowymi? Czy porównywano wektory 180-elementowe opisujące wskaźniki dla każdej z analizowanych par obrazów? Przyjęta podczas porównywania wskaźników jakości obrazów konwencja (1-MSE na str. 73) została chyba naruszona w rozdziale 6.3, kiedy to Autor w punkcie 5 schematu *benchmarka* rekomenduje jako najlepszą „najmniejszą ... wartość wskaźnika jakości".

Reasumując brak możliwości doszukania się przeze mnie odpowiedzi na powyższe wątpliwości/pytania nie umniejsza w żaden sposób dokonań doktoranta i rangi osiągnięć przedstawionych w dysertacji.

***7. Jaka jest przydatność rozprawy do nauk technicznych?***

Zarówno opracowana metoda jak i towarzysząca jej analiza efektywności metod oceny jakości obrazów wykazuje istotną przydatność dla dyscypliny informatyka techniczna i telekomunikacja w dziedzinie nauk inżynieryjno-technicznych. Zarówno zaproponowana metodologia, przygotowana przez Autora baza danych obrazów referencyjnych jak i same wnioski z przeprowadzonych badań istotnie wpływają na obecny stan wiedzy w zakresie metod oceny jakości obrazów cyfrowych. Wzmocnieniem przydatności opracowanych rozwiązań jest zaproponowane w rozdziale 6.3 studium przypadku aplikacji do automatycznej kalibracji ustawień trybów renderingu, przy zachowaniu minimalnych strat jakościowych na rzecz zapewnienia wydajności środowiska renderującego. Ważnym atutem zaproponowanego rozwiązania jest posługiwanie się wyłącznie jednym parametrem, nazwanym przez autora „wskaźnikiem jakości", który zapewnia łatwość procesu automatyzacji oceny jakości obrazów.

***8. Do której z następujących kategorii Recenzent zalicza rozprawę:***

*a/ nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy*
*b/ wymagająca wprowadzenia poprawek i ponownego recenzowania*
*c/ spełniająca wymagania*
*d/ spełniająca wymagania z wyraźnym nadmiarem*
*e/ wybitnie dobra, zasługująca na wyróżnienie*

***9. Konkluzja***

Stwierdzam, że Autor rozprawy wniósł odpowiedni wkład do dziedziny informatyka w dyscyplinie naukowej informatyka techniczna i telekomunikacja. Uważam, że przedłożona do recenzji rozprawa spełnia wymagania Ustawy o stopniach naukowych i tytule naukowym z 14 marca 2003 roku, Dziennik Ustaw Nr 65, poz. 595 (z późniejszymi zmianami). Wnioskuję również o wyróżnienie rozprawy.

Wniosek o wyróżnienie motywuję rangą osiągnięcia naukowego doktoranta – autorskiej metryki oceny jakości obrazów cyfrowych z pełną referencją, którego wartość została udowodniona według najwyższych standardów naukowych obowiązujących w dyscyplinie. Badania zmierzające do opracowania metryki zyskały już widoczność w bazach Scopus, WoS i Google Scholar, a Autor wskazał, w ramach studium przypadków użycia, namacalny scenariusz praktycznego wykorzystania opracowanej metody.

Adam Wojciechowski