

Prof. dr hab. inż. Krzysztof Krawiec
Instytut Informatyki
Politechnika Poznańska
ul. Piotrowo 2
60-965 Poznań

Poznań, 7.06.2025

Recenzja rozprawy doktorskiej mgr inż. Kacpra Kani p.t. "Controllability of Radiance Fields with Sparse Information"

1. Tematyka rozprawy

Przedmiotem rozprawy doktorskiej mgr inż. Kacpra Kani są modele uczenia głębokiego przeznaczone do modelowania i generowania scen w widzeniu komputerowym. Zadaniem stawianym przed algorytmami rozważanymi w rozprawie jest skonstruowanie, na podstawie relatywnie niewielkiej liczby przykładów-obrazów 2D, trójwymiarowego (3D) modelu sceny (statycznego lub dynamicznego) przy pomocy którego można następnie generować (renderować) inne widoki tej samej (lub zbliżonej) sceny. Tematyka ta leży na przecięciu widzenia komputerowego (*computer vision*) i grafiki komputerowej (*computer graphics*), co pozwala mi stwierdzić że praca jednoznacznie wpisuje się w obszar dziedziny Informatyka Techniczna i Telekomunikacja.

2. Ocena treści rozprawy i wkładu oryginalnego

2.1 Treść rozprawy

Rozprawa została przygotowana w języku angielskim, składa się z 6 rozdziałów, i koncentruje się na szeroko rozumianej kontrolowalności rozważanych modeli, co jest umotywowane brakiem tej właściwości w zdecydowanej większości istniejących metod. To niedomaganie jest istotnym wyzwaniem w praktycznych zastosowaniach, zwłaszcza w generatywnych scenariuszach użycia, gdzie intencją użytkownika jest często uzyskanie zamierzonych efektów (np. określonej pozy/ekspresji, oświetlenia, punktu widzenia, czy innej charakterystyki sceny), a zatem niezbędne jest zapewnienie mu odpowiednich do tego środków. W tym sensie rozprawa adresuje istotny problem praktycznej natury.

Punktem wyjścia do proponowanych metod są dwa dominujące obecnie podejścia, tj. Neuronowe Pola Radiancji (ang. Neural Radiance Fields, NeRF) oraz 3D Gaussian Splatting, 3DGS. Autor zdecydował się nazywać to drugie podejście "Rozsmarowywaniem Gaussów", co brzmi nieco

dyskusyjnie w mojej opinii, choć przyznaję że eleganckie przetłumaczenie tego terminu nie należy do łatwych (może np. "Plaskanie Gaussami"?). Proponowane w rozprawie metody rozszerzają te podejścia w nietrywialny sposób i budują nich, adresując ich słabości. Autor bazuje zatem na najnowszych osiągnięciach w tym obszarze badań, co czyni pracę zdecydowanie aktualną.

Rozdziały 2-5 bazują na artykułach opublikowanych lub zaakceptowanych na czołowych konferencjach widzenia komputerowego (m.in. CVPR i ICCV) i opisują główne przyczynki pracy. W rozdziale 2 autor proponuje metodę **CoNeRF**, w której częściowa kontrolowalność sceny została uzyskana poprzez umożliwienie użytkownikowi na nieskomplikowane anotacje niewielkiej liczby obrazów przy pomocy semantycznie bogatych etykiet (orientacyjne wskazanie fragmentu obrazu oraz powiązanej z nim etykiety). W zamian za to metoda udostępnia użytkownikowi kontrolowalne parametry ('suwaki') przy pomocy których może on następnie kontrolować charakterystykę widoku sceny w sensie interpolacji pomiędzy stanami powiązanych z etykietami. Funkcjonalność tę autor osiągnął rozszerzając w nowatorski i interesujący sposób metodę HyperNeRF (Fig. 3) i angażując złożone funkcje straty (Sec. 2.3.2 i kolejne). Zaproponowane podejście skutecznie interpoluje pomiędzy punktami w kontrolowanej przestrzeni, generując nowe widoki sceny (twarze i inne obiekty) o wysokiej, fotorealistycznej jakości, nawet gdy użytkownik przygotowuje anotacje zaledwie kilku obrazów.

Rozdział 3 koncentruje się na modelowaniu obrazu twarzy, które to zadanie jest szczególnie trudnym wyzwaniem w widzeniu komputerowym (choć proponowane podejście da się uogólnić poza to zastosowanie, co autorzy deklarują w Tabeli 5). Autor proponuje metodę **BlendFields** stanowiącą hybrydę podejścia opartego na danych (data-driven) i podejścia opartego na modelu (model-driven), co z jednej strony zapewnia znaczną dozę kontroli nad renderowanymi obrazami, a z drugiej poprawia jakość generowanych obrazów (w porównaniu z podejściem opartym wyłącznie na modelu). Celem szczegółowym proponowanego podejścia jest fotorealistyczne renderowanie detali, w szczególności zmarszczek. Zamierzony efekt uzyskany jest przez uzależnienie radiancji (radiance) od ekspresji (Eq. 27) przy pomocy zestawu szablonów deformacji, i korygowanie bazowego modelu (Fig. 18), co stanowi oryginalny przyczynek tego rozdziału (i powiązanej z nim publikacji). Tak wprowadzone korekty są dodatkowo 'bramkowane' na bazie lokalnego podobieństwa siatki czworokątów do siatki referencyjnej (sekcja 3.3.2), związanej z daną ekspresją twarzy, a następnie mieszane liniowo (blending), przy jednoczesnym zapewnieniu gładkości odwzorowania (sekcja 3.3.3). Autor ilustruje także możliwość zastosowania proponowanego podejścia do obiektów innych niż twarz (Fig. 22).

Rozdział 4 bazuje na drugiej z wymienionych wcześniej technik, tj. Gaussian Splatting, i opisuje propozycję metody **LumiGauss** zorientowanej na rekonstrukcję i kontrolowaną syntezę pól radiancji, poprzez ich uprzednią 'faktoryzację' z obrazów scen. Gaussian Splatting jest tu wykorzystane do konstrukcji potoku 'grafiki odwrotnej' (*inverse graphics*) i zamodelowania obrazu w postaci zbioru dwuwymiarowych rozkładów Gaussa, parametryzowanych (w uczalny sposób) wektorami normalnymi, albedo i funkcjami transferu. Otrzymany w ten sposób parametryczny opis sceny może być następnie wykorzystany do tworzenia jej alternatywnych renderingów z wykorzystaniem innych, kontrolowalnych modeli oświetlenia. Podejście bazuje na fizycznie

uzasadnionym i znanym z grafiki komputerowej modelu transferu radiancji, i co m.in. umożliwia operowanie w dwóch trybach, z i bez modelowania cieni. Autor zilustrował skuteczność metody na zadaniach rekonstrukcji sceny, rekonstrukcji ceny obserwowanej z innego punktu, zmianie położenia/charakterystyki oświetlenia (relighting), predykcji cienia, albedo i wektorów normalnych. Zarówno zestawienia jakościowe (wizualizacje) jak i porównania ilościowe na kilku relewantnych metrykach (PSNR, MSE, MAE, SSIM; Tabela 10) wskazują na zalety LumiGauss w odniesieniu do (niemal dziesięciu) algorytmów referencyjnych.

Rozdział 5 prezentuje ostatni z głównych przyczynków rozprawy, metodę **CLoG**, której przyświeca optymalizacja wydajności kontrolowalnego renderingu w oparciu o Gaussian Splatting, które to podejście w podstawowym wydaniu jest kosztowne obliczeniowo. Autor proponuje osiągnąć ten cel poprzez dynamiczne dostosowanie liczby stosowanych rozkładów Gaussa ('splats') do zadanego poziomu szczegółowości/detali. Istotną zaletą podejścia jest możliwość kontrolowania tego poziomu przy pomocy *ciągłego* parametru (level of detail, LoD, factor l) w miejsce kontroli skokowej/dyskretnej, typowej dla wcześniej proponowanych metod. Autor uzyskał to poprzez nauczanie modelu neuronowego realizującego dwuwymiarową ciągłą mapę/odwzorowanie, umożliwiającego wydajne próbkowanie na dowolnym poziomie precyzji/rozdzielczości (mimo tego że model pozyskiwany jest na jednej rozdzielczości). Ciekawym i nowatorskim elementem metody jest m.in. wieloetapowa procedura uczenia modelu (5.3.3–5.3.5), która z drugiej strony jest też dość skomplikowana. Autor przeprowadził ocenę metody CLoG na szeregu benchmarków (modelowanie widoków twarzy i innych obiektów), dowodząc jej atrakcyjności w sensie nie tylko poszczególnych metryk, ale przetargu pomiędzy jakością renderingu a liczbą Gaussów (m.in. Fig. 43).

Rozdział 6 podsumowuje rozprawę, dyskutuje jej przyczynki i oczekiwany impakt, oraz wyznacza kierunki dalszych prac.

2.2 Ocena wkładu oryginalnego i prezentacji pracy

Praca jest bardzo zwarta tematycznie, a proponowane algorytmy prezentowane są na odpowiednim poziomie szczegółowości, pozwalając na zrozumienie zaproponowanych mechanizmów z jednej strony, a z drugiej unikając zbyt detalicznych formalizmów. Autor umiejętnie buduje na istniejących metodach i obserwacjach dotyczących ich architektury i efektów działania, adresując ich istotne słabości. Eksperymenty dotyczą wielu zbiorów danych (rzeczywistych i syntetycznych), mają wieloaspektowy charakter (w sensie badania różnych charakterystyk algorytmów i produkowanych przez nich wyników) i przeprowadzone zostały bardzo rzetelnie. Proponowane algorytmy zestawiane są z czołówką konkurencyjnych metod i w większości przypadków okazują się je przewyższać na rozważanych kryteriach, lub nie odbiegać od nich znacząco. Wyniki prezentowane są zarówno przez przykłady/wizualizacje jak i, tam gdzie to możliwe, także ilościowo, przy pomocy metryk. Autor umiejętnie wykorzystuje m.in. ablacje, tj. uproszczone/zredukowane warianty proponowanych algorytmów, aby wykazać istotność wybranych komponentów architektur. Co szczególnie doceniam, to autor jest też świadom ograniczeń proponowanych podejść, analizuje przypadki w których one zawiodły (np. Sec. 3.4.4) i dyskutuje je [w każdym rozdziale. Badania opisane w rozprawie są dobrze umotywowane licznymi zastosowaniami praktycznymi w widzeniu

komputerowym i grafice komputerowej. Pytania badawcze (hipotezy) sformułowane w Sekcji 1.2 zostały pozytywnie zweryfikowane.

Odnosnie aspektów które personalnie uważam za szczególnie ciekawe, proponowane podejścia stanowią interesującą demonstrację użyteczności 'liftingu' jako komponentu w złożonych architekturach głębokiego uczenia (m.in. w Rozdziale 2). Łączą też w nietrywialny sposób aspekty 'rastrowe' i 'geometryczne' obrazów i scen (np. wykorzystanie zmian w objętości czworościanów siatki w Rozdziale 3, sekcja 3.3.2). W szerszym kontekście, badania nad metodami uczenia maszynowego które modelują sceny (a nie jedynie je analizują) uważam za szczególnie pożądane i przyszłościowe, jako że wymagają one od systemu uczącego się właściwego 'zrozumienia' zawartości sceny (a czasami także zachodzących w niej procesów), a to z kolei jest warunkiem koniecznym dla zapewnienia wyjaśnialności i interpretowalności podejmowanych decyzji (nie mówiąc już o innych pozytywnych implikacjach, jak np. redukcja zapotrzebowania na etykietowane dane, co dr Kania wykazał m.in. w Rozdziale 2).

Redakcja pracy jest zasadniczo nienaganna; szczególnie doceniam zwięzłość przy zachowaniu jednocześnie wysokiej precyzji notacji i opisu. W związku z tym mam jedynie nieliczne uwagi redaktorskie. Okazjonalnie zdarza się że uproszczenia/skróty notacyjne prowadzą do drobnych niekonsekwencji. Sygnatury odwzorowań, stosowane m.in. w sekcji 2.3.3, byłyby pomocne także w innych miejscach, np. w niektórych wcześniejszych równaniach tego samego rozdziału. Niektóre diagramy ilustrujące poszczególne komponenty mogłyby bardziej jednoznacznie odwoływać się do symboli w notacji formalnej (np. Fig. 4). Lektura rozdziału 5 byłaby łatwiejsza gdyby znalazł się w nim, podobnie jak w pozostałych rozdziałach, diagram wyjaśniający etapy działania metody.

Z uwag polemicznych, metody proponowane w rozprawie (a przynajmniej niektóre z nich, m.in. np. CLoG, Sec. 5.3.3 i 5.3.4) wydają się wymagać dużego nakładu pracy badacza w procesie uczenia (tzw. model babysitting). Nasuwa to pewne wątpliwości odnośnie odporności/stabilności proponowanych algorytmów i podejrzenie że zaaplikowanie ich do zbiorów danych/problemów o innej naturze wymagałoby ich odmiennej parametryzacji i kolejnych znacznych nakładów czasu. Zastanawiam się także nad zdolnościami proponowanych metod do ekstrapolacji. Modele głębokie słyną ze znakomitych zdolności interpolowania *pomiędzy* obserwacjami, ale często zawodzą gdy próbujemy zmusić je do ekstrapolacji *poza* 'domenę' zbioru uczącego. Czy np. CoNeRF potrafi też ekstrapolować, raczej niż jedynie interpolować? (np. pomiędzy położeniami, orientacjami obiektów, ale także pomiędzy wartościami kontrolującymi atrybutów) Termin 'ekstrapolacja' pojawiają się w m.in. w RQ2 i Rozdziale 3, ale nie doczekał się chyba w rozprawie definitywnych konkluzji.

Uwagi te mają jednak stanowić przydatną informację zwrotną i nie wpływają na moją ogólnie pozytywną ocenę przedłożonej rozprawy.

Mgr Kania jest pierwszym autorem trzech z czterech prac stanowiących podstawę rozdziałów rozprawy, co sugeruje jego esencjonalny wkład. Prace te prezentowane były na czołowych konferencjach widzenia komputerowego (m.in. CVPR i ICCV), co jest dodatkowym probierzem ich

jakości. Spośród pozostałych publikacji mgr Kani przynajmniej jedna ukazała się też w czasopiśmie (Sec. 1.5, Pattern Recognition Letters).

3. Konkluzja końcowa

Rozprawa doktorska mgr inż. Kacpra Kani zawiera szereg oryginalnych osiągnięć i wyników, unikalnych w skali globalnej i dotyczących aktualnych tematów badawczych o znaczącym przełożeniu praktycznym. Uważam że hipotezy postawione przez Autora pracy zostały zdecydowanie potwierdzone. **Moja ocena rozprawy jest zatem zdecydowanie pozytywna.**

Wobec powyższego stwierdzam, że **rozprawa doktorska mgr inż. Kacpra Kani spełnia z solidną nawiązką warunki stawiane przez ustawę o tytule naukowym i stopniach naukowych w odniesieniu do rozpraw doktorskich, a zatem powinna być dopuszczona do publicznej obrony, o co wnoszę do Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej.**

Ponadto, biorąc pod uwagę nowatorski charakter przyczynków, skuteczne wykazanie konkurencyjności proponowanych metod względem istniejących algorytmów, znaczne przełożenie praktyczne wyników, oraz publikacje autora w materiałach czołowych konferencji i czasopismach, **wnioskuję o wyróżnienie rozprawy.**



PODPIS ZAUFANY

KRZYSZTOF
KRAWIEC

07.06.2025 10:42:31 (GMT+2)

Dokument podpisany elektronicznie
podpisem zaufanym