RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ
zaprasza na
OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr. inż. Kacpra Kani

która odbędzie się w dniu **30.09.2025 roku,** o godzinie **12:00** w trybie stacjonarnym

Temat rozprawy:

Controllability of Radiance Fields with Sparse Information

Promotor: prof. dr hab. inż. Tomasz Trzciński – Politechnika Warszawska

Promotor pomocniczy: dr inż. Marek Kowalski

Recenzenci:

prof. dr hab. inż. Bogusław Cyganek - Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie

prof. dr hab. inż. Krzysztof Krawiec – Politechnika Poznańska

prof. dr Marc Pollefeys - Politechnika Federalna w Zurychu

Obrona odbędzie się w formie stacjonarnej w sali nr 40 w Gmachu Wydziału Matematyki i Nauk Informacyjnych Politechniki Warszawskiej. Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1, oraz na stronie internetowej BIP PW: www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techniczna-i-Telekomunikacja/mgr-inz.-Kacper-Kania

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
Prof. dr hab. inż. Jarosław Arabas

Streszczenie

Rozprawa doktorska obejmuje szereg publikacji, które wprowadzają nowatorskie metody renderowania postaci ludzkich, używających częściowej informacji. Zaproponowane podejścia skupiają się wokół Neuronowych Pól Radiancji (z ang. *Neural Radiance Fields* lub NeRFs) oraz Rozsmarowywania 3D Gaussów (*3D Gaussian Splatting* lub 3DGS). Przedstawiamy jak te modele budują reprezentację 3D ze zdjęć 2D oraz demonstrujemy podejścia, za pomocą których możemy warunkować te reprezentacje w celu uzyskania wysokiej jakości animacji postaci ludzkich i innych obiektów przedstawionych na scenie. Prezentujemy techniki, które używają proste oraz interpretowalne wejścia wytworzone z rzadkich danych trenujących. Pokazujemy również, jak można rozszerzyć te podejścia do uczenia w ujęciu kilkustrzałowym (*few-shot learning*).

Rozprawę zaczynamy od rozważań na temat neuronowych pól radiancji, adresującym przy tym ograniczenia istniejących podejść, oraz od kontrybucji w dziedzinie kontrolowalnych pól radiancji. Poprzez integrację częściowej oraz rzadkiej informacji w czasie uczenia, nasze rozszerzenie NeRFów—CoNeRF—wykorzystuje gładkość sieci neuronowych, aby predykować kontrolowalne, wysokiej jakości obrazy postaci ludzkich oraz obiektów.

Aby ograniczyć zależność istniejących metod od ekstensywnych, wysokiej jakości adnotacji w nagraniach wideo z wielu ujęć jednoczesne, wprowadzamy nową metodę, BlendFields, uczenia NeRFów w ujęciu kilkustrzałowym. Nasze podejście uczy się wewnętrznych szablonów deformacji obiektu, które są mieszane w ciągły i gładki sposób w czasie wnioskowania, tym samym osiągając znaczącą poprawę jakości rekonstrukcji obrazu w porównaniu do istniejących podejść. Nasze podejście pozwala na efektywne i realistyczne renderowanie ludzi z zaledwie kilku statycznych ujęć.

W tej rozprawie adresujemy również brak metod, które pozwalą na adaptację modeli pól radiancji dla określonych zdolność obliczeniowych w czasie wnioskowania. Przedstawiamy podejście do uczenia „od-ogółu-do-szczegółu” dla modeli 3DGS o nazwie CLoG, które dynamicznie powiększa przestrzennie siatkę ukrytych, wyuczalnych wektorów, kodujących informacje na temat Gaussów w przestrzeni 3D. Ta strategia osiąga porównywalne wyniki do podstawowego modelu 3DGS, jednocześne pozwalając na wdrożenie tej reprezentacji do urządzeń o różnych zdolnościach obliczeniowych przy pojedynczym uczeniu modelu.

Dodatkowo, proponujemy podejście do kontrolowania światła środowiskowego w polach radiancji przy użyciu naszego podejścia o nazwie LumiGauss. Poprzez uwzględnienie wyliczonej funkcji przeniesienia radiancji, model ten pozwala na zmianę oświetlenia sceny w realistyczny, oparty na zjawiskich fizycznych sposób.

Przedstawione badania rozwijają naszą wiedzę w zakresie kontrolowania pól radiancji oraz rozszerzają ją w zakresie aplikacji do środowisk, w których tylko kilka próbek uczących jest dostępnych. Te innowacje zwiększają możliwości, w jakich możemy rekonstruować postacie ludzkie oraz codzienne obiekty przedstawione na scenie w 3D z ograniczonych danych oraz otwierają nowe ścieżki badań w tej dziedzinie.

Słowa kluczowe: Renderowanie Sieciami Neuronowymi, Neuronowe Pola Radiancji, Uczenie Kilkustrzałowe, Renderowanie Postaci Ludzkich, Warunkowanie Częściową Informacją, Rozsmarowywanie Gaussów

Prof. dr hab. inż. Bogusław Cyganek
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie
Wydział Informatyki, Elektroniki i Telekomunikacji
Instytut Elektroniki
cyganek@agh.edu.pl

Kraków, 20.05.2025

Rada Naukowa Dyscypliny
INFORMATYKA TECHNICZNA
I TELEKOMUNIKACJA
Sekretariat
Data wpływu 27.05.25r.
Numer.....

Recenzja rozprawy doktorskiej
Pana magistra inżyniera Kacpra Kani
zatytułowanej:
Controllability of Radiance Fields with Sparse Information

Wstęp

Recenzja dotyczy rozprawy doktorskiej Pana magistra inżyniera Kacpra Kani pt. „*Controllability of Radiance Fields with Sparse Information*", która powstała w roku 2025 w Politechnice Warszawskiej. Promotorem pracy jest Pan prof. dr hab. inż. Tomasz Trzciński, a promotorem pomocniczym – Pan dr Marek Kowalski. Przygotowanie recenzji zostało wykonane na zlecenie Przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja, Pana prof. dra hab. inż. Jarosława Arabasa. Rozprawa doktorska została napisana w języku angielskim. Recenzja sporządzona została w postaci odpowiedzi na pytania dotyczące rozprawy doktorskiej Pana mgra inż. Kacpra Kani.

1. Problem badawczy i jego znaczenie

Jaki jest najważniejszy problem rozważany w rozprawie?

Żyjemy w czasach wszechobecnej grafiki komputerowej, której podstawy działania opierają się głównie na kreowaniu fotorealistycznych obrazów. Jednakże uzyskanie realistycznie wyglądających obrazów odzwierciedlających strukturę złożonych scen nie należy do procesów ani prostych ani też łatwych obliczeniowo. Dlatego od lat trwają badania nad opracowaniem coraz to lepszych metod, jak również platform komputerowych oraz kart graficznych o coraz większych mocach obliczeniowych. Komputerowo generowane obrazy syntetyczne są otrzymywanie w wyniku działania tzw. algorytmów renderowania obrazów. Tradycyjne podejście opiera się tu o metody rasteryzacji lub ray-tracingu, które jako dane wejściowe przyjmują odpowiednio zdefiniowane reprezentacje sceny w postaci parametrów geometrii scen, obiektów, parametrów kamery, jak również właściwości materiałów i oświetlenia. Dla przykładu, sceny mogą być dobrze aproksymowane za pomocą tzw. siatki triangulacyjnej (ang. *triangle meshes*) wraz z teksturami, siatek wolumetrycznych (ang. *volumetric grids*), chmur punktów 3D (ang. *point clouds*), czy też uwikłanych funkcji powierzchniowych (ang. *implicite surface functions*). Zdarza się jednak, że znane są jedynie obrazy z wybranych różnych punktów obserwacji sceny, natomiast pozostałe parametry fizyczne muszą być w jakiś sposób oszacowane. Jest to zagadnienie – zwane renderowaniem odwrotnym – które jest niezwykle trudne i kosztowne obliczeniowo, szczególnie jeśli chcemy otrzymać realistycznie wyglądające obrazy. Zagadnieniem ściśle spokrewnionym jest problem syntezy nowych obrazów na podstawie obrazów uzyskanych w innych punktach widzenia kamer,

a możliwe że też w różnych chwilach czasowych. Dzięki takiej syntezie możliwe jest uzyskanie widoku z miejsca, w którym kamera nie była umieszczona, bądź też odtworzenie brakujących klatek np. w uszkodzonym filmie. Tradycyjne podejście to rekonstrukcja oraz reprojekcja przestrzeni 3D na podstawie pary obrazów stereoskopowych (ang. *stereovision*), bądź też układu z więcej niż dwoma obrazami. Innym podejściem jest uzyskanie informacji o strukturze przestrzeni 3D na podstawie strumienia obrazów, czyli np. z wideo – jest to tzw. struktura-z-ruchu (ang. *Structure-from-Motion* – SfM). Metody te, opracowywane od lat, dobrze sobie radzą z prostymi scenami sztywnych obiektów. Jednakże często zawodzą w przypadku scen dynamicznych, czy też w przypadku ograniczonego obszaru widzenia kamer. Stąd też w ostatnich latach powstały metody nowsze, oparte na algorytmach sztucznej inteligencji (ang. *Artificial Intelligence* – AI).

Jednym z najnowszych podejść w tej dziedzinie są algorytmy renderowania z wykorzystaniem neuronowych pól radiancji (ang. *Neural Radiance Field* – NeRF). Bazują one na nowoczesnych architekturach głębokich sieci neuronowych, które generalnie rzecz biorąc dokonują transformacji informacji opisującej sceny, podanych na przykład w postaci współrzędnych punktów oraz kątów kierunkowych widzenia, w dane wyjściowe w postaci radiancji w danym punkcie przestrzeni. Przewaga metod renderowania neuronowego nad podejściami tradycyjnymi polega na możliwości wytrenowania wszystkich parametrów sceny na podstawie obrazów treningowych, tj. dostępnych obrazów z wybranych punktów obserwacji, jednakże bez konieczności jawnego obliczania parametrów wspomnianej transformacji renderowania odwrotnego.

Tym niemniej, wyżej opisane metody bazujące na głębokich sieciach neuronowych mają też pewne wady. Jedną z głównych jest olbrzymia złożoność obliczeniowa, szczególnie w procesie trenowania, który również wymaga dostępności znaczącej liczby danych trenujących. Stąd też wiele prac usprawniających ten proces, jak również szukających nowych i bardziej dogodnych dróg. Ciekawym i w wielu aspektach alternatywnym rozwiązaniem są metody reprezentacji scen bazujące na wielowymiarowych funkcjach gaussowskich do reprezentacji radiancji wolumetrycznej w renderowanej scenie. Parametry tych funkcji wyznaczane są między innymi na podstawie rzadkich punktów otrzymanych w procesie kalibracji kamer, często za pomocą metody SfM, a sam proces obliczeniowy jest znacznie lżejszy obliczeniowo niż w przypadku rozwiązań wymagających trenowania głębokich sieci neuronowych. Należy dodać, że sam proces renderowania sceny jest dodatkowo uzupełniany przez tzw. anizotropowe rozmazywanie (ang. *anisotropic splatting*).

Praca naukowa oraz metody zaprezentowane w rozprawie doktorskiej Pana magistra inżyniera Kacpra Kani wpisują się w wyżej opisane obszary badawcze związane z opracowaniem nowoczesnych metod grafiki komputerowej. Główne obszary badawcze podjęte przez Doktoranta to:

1. Opracowanie ulepszonej wersji algorytmu renderowania z wykorzystaniem neuronowych pól radiancji w postaci metody CoNeRF, której główną zaletą jest możliwość niezależnej kontroli wyspecyfikowanych fragmentów syntezowanej sceny.
2. Opracowanie metody BlendFields do modelowania wybranych deformacji w obrazach, takich jak marszczenie twarzy, za pomocą kilku ujęć z wielu kamer.
3. Opracowanie metody LumiGauss do kontrolowania pól radiancji poprzez oświetlenie środowiskowe dla ujęć naturalnych otrzymywanych przy różnych parametrach oświetlenia.
4. Opracowanie metody CLoG do redukcji złożoności obliczeniowej przy renderowaniu scen z wykorzystaniem funkcji gaussowskich 3D oraz mechanizmu rozmazywania anizotropowego.

Wyżej wymienione zagadnienia wprost odpowiadają tezom pracy przedstawionym w Rozdziale 1 dysertacji.

Zagadnienia naukowe podjęte przez Pana magistra inżyniera Kacpra Kanię mają duże znaczenie zarówno w aspekcie teoretycznym, jak i eksperymentalnym. Opracowane oryginalne metody generowania obrazów dla celów grafiki komputerowej mogą znaleźć szerokie zastosowanie we wszelkich systemach związanych z grafiką komputerową, takich jak m.in. przemysł filmowy, systemy wirtualnej rzeczywistości (ang. *virtual reality*), gry komputerowe itd. Stąd też podjęty problem badawczy oceniam jako bardzo znaczący, a sama praca ma charakter teoretyczno-doświadczalny.

Czy praca ma charakter naukowy?

Problemy podjęte w dysertacji doktorskiej Pana magistra inżyniera Kacpra Kani mają charakter naukowy – w światowym środowisku naukowym ciągle trwają prace badawcze i rozwojowe w tym kierunku, jak również są publikowane nowe metody przetwarzania obrazów. Tym niemniej, ciągle nie istnieją algorytmy i metody, które w pełni spełniałyby wymagane założenia. Dlatego istnieje potrzeba kontynuacji badań w tym zakresie, a dostarczenie nowych rozwiązań w dobie dominacji metod AI wymaga przede wszystkim opracowania nowych podejść i oryginalnych zbiorów danych. Niemniej ważnym jest sam problem złożoności obliczeniowej tej grupy metod – tu też trwają liczne badania nad jej redukcją. Niezwykle istotne jest też opracowanie skutecznych metod ich analizy i oceny, które uwzględniają skomplikowane zależności występujące w warunkach rzeczywistych. Zarówno zagadnienia naukowe, jak i wyzwania eksperymentalne oraz sposoby ich rozwiązania zostały zaprezentowane w rozprawie doktorskiej Pana mgra inż. Kacpra Kani.

Czy praca ma znaczenie praktyczne?

Podjęty przez Pana mgra inż. Kacpra Kani problem badawczy niewątpliwie posiada istotny walor i znaczenie praktyczne – zaprezentowane rozwiązania mogą wprost prowadzić do ich aplikacji we wszelkiego rodzaju systemach związanych z grafiką komputerową, czyli takich jak generowanie filmów, gry komputerowe, systemy wirtualnej rzeczywistości i wiele innych.

5. Wkład Autora

Jaki jest najważniejszy wkład Autora rozprawy?

Poniżej zostały przedstawione główne osiągnięcia naukowe Pana magistra inżyniera Kacpra Kani zeprezentowane w Jego rozprawie doktorskiej.

1. *Opracowanie ulepszonej wersji algorytmu renderowania z wykorzystaniem neuronowych pól radiancji w postaci metody CoNeRF, której główną zaletą jest możliwość niezależnej kontroli wyspecyfikowanych fragmentów syntezowanej sceny.*

   Metoda CoNeRF (ang. *Controllable* NeRF) rozszerza znany algorytm syntezy scen z wykorzystaniem neuronowych pól radiancji NeRF. Tak jak już wspomniano, metody tego typu wykorzystują wielowarstwowe głębokie sieci neuronowe o pełnych połączeniach (ang. *multi-layer perceptron* – MLP), które dokonują transformacji poprzez próbkowanie pięciu współrzędnych – 3 przestrzenne oraz 2 związane z kierunkami widzenia wzdłuż promieni kamer – w wyjściową wartość koloru oraz współczynnika gęstości w danej jednostce wolumetrycznej. Przewaga metod renderowania neuronowego nad podejściami tradycyjnymi polega na możliwości wytrenowania wszystkich parametrów sceny na podstawie obrazów

treningowych, tj. obrazów dostępnych z wybranych punktów obserwacji, jednakże bez konieczności jawnego obliczania parametrów wspomnianej transformacji renderowania odwrotnego. Jednakże tradycyjny NeRF nie umożliwia precyzyjnej kontroli, czy też wpływania na wygląd wybranych fragmentów renderowanej sceny, a nie całej sceny jako całości. Jest to ogólniejszy problem w złożonych systemach AI, np. jak wpływać na szczegóły rozwiązania za pomocą cokolwiek krótkiego i ogólnego prompta. W przypadku NeRF kontrola tego typu byłaby wielce użyteczna, np. dla grafików którzy pragną zmienić wygląd lewego oka, czy np. układu ust tworzonego obrazu. Pan mgr inż. Kacper Kania zrealizował to zadanie poprzez zaproponowanie oryginalnej metody CoNeRF. Główny pomysł to zaproponowanie metody, która wpływałaby na wybrane wartości wag w części semantycznej sieci, tj. te wyłącznie odpowiedzialne i mające wpływ na indywidualne cechy obrazu. W tym celu CoNeRF wymaga poetykietowania przez użytkownika, w niewielkiej liczbie obrazów treningowych, wybranego za pomocą maski miejsca, czy też miejsc, na które chcemy wpływać. Ale główna nowość polega na potraktowaniu atrybutów obrazu, takich jak np. wyraz twarzy, jako zmiennych ukrytych (ang. *latent variables*), które można wyznaczyć za pomocą sieci neuronowej i procesu regresji. W ten sposób atrybuty mogą zostać „odkryte" przez sieć w sposób automatyczny dla innych obrazów, tj. nie używanych do trenowania. Jest to tzw. uczenie z zaledwie kilkoma przykładami (ang. *few-shot learning*).

Główne zalety metody CoNeRF opracowanej przez Pana mgra inż. Kacpra Kanię można podsumować następująco:

- Umożliwia ona bezpośrednią i precyzyjną kontrolę nad reprezentacjami wybranych cech i fragmentów obrazowych w reprezentacji NeRF.
- Do trenowania wystarcza niewielka liczba poetykietowanych obrazów w formie wartości atrybutów i odpowiadającej im maski 2D.
- Przetestowana w aplikacji animacji twarzy wykazuje lepsze właściwości wizualne niż inne konkurencyjne metody, takie jak np. HyperNeRF.
- Metoda może znaleźć liczne aplikacje w praktycznych zadaniach grafiki komputerowej.

Praca ta została wykonana przez p. mgra inż. Kacpra Kanię przy współpracy z zespołem naukowym. Została ona opublikowana na prestiżowej konferencji IEEE Computer Vision and Pattern Recognition (CVPR) 2022. Pan mgr inż. Kacper Kania jest pierwszym z autorów.

Kania, K., Yi, K. M., Kowalski, M., Trzciński, T., and Tagliasacchi, A., "*CoNeRF: Controllable Neural Radiance Fields*", CVPR, 2022.

2. *Opracowanie metody BlendFields do modelowania wybranych deformacji w obrazach, takich jak marszczenie twarzy, za pomocą kilku ujęć z wielu kamer.*

Chyba jednym z najbardziej prestiżowych zadań w grafice komputerowej jest generowanie realistycznie wyglądających ludzkich twarzy. Jest to jednak zadanie trudne, a istniejące metody bazują na trenowaniu z wykorzystaniem wielkich repozytoriów danych. Są też metody prostsze, które jednak nie są w stanie uchwycić szczegółów twarzy, jej właściwej ekspresji, czy też niuansów wyrazu twarzy. Metoda BlendFields została opracowany przez Pana mgra inż. Kacpra Kanię aby wypełnić tę lukę. W swoim zamierzeniu służy ona właśnie do renderowania szczegółów twarzy zależnych od ekspresji, co odpowiada składowym wysokoczęstotliwościowym, podczas gdy sieci neuronowe mają naturalną tendencję do uwypuklania roli składowych niskoczęstotliwościowych. Ekspresje twarzy, które nie były dostępne podczas uczenia są modelowane poprzez odpowiednie miksowanie ekspresji

skrajnych z dostępnego zbioru. Mieszanie jest sterowane dzięki pomiarowi lokalnych zmian w polu wolumetrycznym wyrazów twarzy. Cenną zaletą tej metody jest możliwość generowania ekspresji twarzy, których nie było w procesie trenowania.

BlendFields czerpie inspirację z klasycznej grafiki komputerowej poprzez mieszanie wyrażeń z danych treningowych w celu renderowania wyrażeń niewidocznych podczas treningu, a częściowo bazuje na wcześniej opisanym podejściu NeRF. Zaletą BlendFields jest możliwość wytrenowania z użyciem relatywnie niewielkiego zbioru obrazów treningowych. Jest to duża zaleta, zarówno ze względu na trudności w dostępie do tego typu obrazów, jak również ze względu na wymaganą moc obliczeniową i czas takiego treningu. Implementacja Pana mgra inż. Kacpra Kani została oparta na metodzie VolTeMorph, która używa wolumetrycznego modelu twarzy 3DMM. BlendFields została zweryfikowana eksperymentalnie z wykorzystaniem między innymi zbioru Multiface, osiągając najlepsze wyniki wśród metod konkurencyjnych, takich jak NeRF, HyperNeRF oraz VolTeMorph, co można zaobserwować choćby poprzez analizę zmarszczek na twarzach.

Podsumowując – główne zalety BlendFields to:

- Renderowanie obrazów w sposób kontrolowany i możliwy do zinterpretowania.
- Możliwość zastosowania do renderowania awatarów ludzkich uczonych z publicznie dostępnych zestawów danych.
- Mimo iż zaprojektowana i zbadania w zadaniu renderowania twarzy, BlendFields może być wykorzystania do renderowania innych obiektów pod warunkiem dostępności danych trenujących.

Praca ta została wykonana przez Doktoranta przy współpracy z zespołem naukowym. Została ona opublikowana na prestiżowej konferencji IEEE CVPR 2023. Pan mgr inż. Kacper Kania jest pierwszym z autorów.

Kania, K., Garbin, S. J., Tagliasacchi, A., Estellers, V., Yi, K. M., Valentin, J., Trzciński, T., and Kowalski, M., "*BlendFields: Few-Shot Example-Driven Facial Modeling*" in CVPR, 2023, pp. 404–415.

3. *Opracowanie metody LumiGauss do kontrolowania pól radiancji poprzez oświetlenie środowiskowe dla ujęć naturalnych otrzymywanych przy różnych parametrach oświetlenia.*

Kolejnym z problemów grafiki komputerowej jest konieczność zmiany parametrów oświetlenia, czy też dopasowania kolorów scen już istniejących. Problem tego typu powstaje w wielu zadaniach, np. podczas kręcenia kolejnych scen w różnym czasie, a następnie post-produkcji. Jest to kolejny bardzo trudny problem, gdyż oddzielenie parametrów oświetlenie sceny, tak aby można było nałożyć inny sposób oświetlenia, jest procesem złożonym obliczeniowo. W celu rozwiązania tego problemu Pan mgr inż. Kacper Kania wraz z zespołem zaproponowali więc metodę o nazwie LumiGauss. Służy ona do rekonstrukcji scen 3D z nałożeniem nowego typu oświetlenia, a wykorzystuje do tego wielowymiarowe funkcje Gaussa wraz z techniką rozmazywania. Zaproponowane podejście umożliwia otrzymywanie scen o wysokiej jakości rekonstrukcji oraz realistycznej syntezie oświetlenia na podstawie tzw. map otoczenia, które – w odróżnieniu od wyodrębnionych specyficznych kolorów obiektów widocznych w scenie – są rodzajem ogólnej tekstury wyodrębnionej dla całej sceny. Metoda LumiGauss umożliwia ponadto poprawę jakości cieni, np. widocznych w scenach plenerowych. Realizowane jest to przez wykorzystanie właściwości funkcji zespolonych tzw. harmonicznych sferycznych. LumiGauss może syntetyzować realistyczne obrazy dla niewidocznych map

otoczenia. Metoda ta została przetestowana przez Pana mgra inż. Kacpra Kanię na zestawie danych NeRF-OSR, wykazując lepszą wydajność w porównaniu z metodami bazowymi.
Oryginalnym podejściem w zaproponowanej metodzie LumiGauss jest wykorzystanie dwuwymiarowych funkcji gaussowskich z rozmazywaniem (ang. 2DGS – *Gaussian Splatting*). Tak jak wspomniano, zastosowanie wielowymiarowych funkcjach gaussowskich do reprezentacji radiancji wolumetrycznej w renderowanej scenie rozwiązuje szereg problemów związanych z wymogami i czasem trenowania, a występujących w przypadku metod bazujących na głębokich sieci neuronowych NeRF. Do predykcji ogólnej mapy otoczenia Pan mgr inż. Kacper Kania ponownie zaproponował użycie sieci typu MLP składającej się z trzech warstw. Między innymi dzięki zastosowaniu funkcji gaussowskich i płytkich sieci neuronowych, w przypadku LumiGauss czas uczenia jest co najmniej o rząd wielkości mniejszy niż w przypadku metod konkurencyjnych, takich jak wspomniana już metoda NeRF-OSR, czy też NeuSky – obydwie zaproponowanych przez innych autorów.
Główne cechy zaproponowanej metody LumiGauss to:

- Wykorzystanie dwuwymiarowych funkcji gaussowskich z rozmazywaniem do tzw. odwrotnego potoku graficznego w warunkach rzeczywistych, co umożliwia uzyskanie wysokiej jakości mapy albedo (kolorów obiektów) oraz uogólnionej mapy otoczenia sceny.
- W celu lepszej reprezentacji cieni w scenach uczone są funkcje transferu promieniowania dla każdego „rozmycia" dwuwymiarowego, które jest za pomocą harmonicznych sferycznych.
- Otrzymane mapy otoczenia mogą być z powodzeniem używane do ponownego oświetlenia dowolnych obiektów, co ma zastosowanie np. w silnikach graficznych gier komputerowych.

Praca ta została wykonana przez Doktoranta przy współpracy z zespołem naukowym. Została ona opublikowana na konferencji IEEE/CVF Winter Conference on Applications of Computer Vision – WACV 2025, Tucson, Arizona USA. Pan mgr inż. Kacper Kania jest drugim z jej autorów.

Kaleta, J., **Kania, K.**, Trzcinski, T., and Kowalski, M., "*LumiGauss: High-Fidelity Outdoor Relighting with 2D Gaussian Splatting*", IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2025.

4. *Opracowanie metody CLoG do redukcji złożoności obliczeniowej przy renderowaniu scen z wykorzystaniem funkcji gaussowskich 3D oraz mechanizmu rozmazywania anizotropowego.*

Celem opracowanej metody o nazwie CLoG (ang. *Continuous Levels of Gaussians*) jest dynamiczna redukcja liczby wielowymiarowych funkcji gaussowskich niezbędnych do prawidłowego renderowania danej sceny. Osiągnięto to dzięki oryginalnemu pomysłowi wykorzystania wyłącznie pojedynczej reprezentacji tego typu funkcji, których parametry wyznaczane są w wyniku jednorazowego trenowania stosunkowo niewielkiej sieci neuronowej typu MLP. Trenowanie to prowadzi do wyznaczenia mapowania informacji pochodzącej z wielokanałowego obrazu 2D na parametry trójwymiarowych funkcji gaussowskich 3D. Ważną cechą tego rozwiązania jest możliwość skalowania do mniejszych rozdzielczości, które również renderowane są prawidłowo z wykorzystaniem mniejszej liczby funkcji gaussowskich.
Główne cechy zaproponowanej metody CLoG to:

- Zaproponowanie oryginalnej techniki ciągłych poziomów funkcji Gaussa (*Continuous Levels of Gaussians* – CLoG), która umożliwia otrzymywanie ciągłych poziomów szczegółowości dla 3D Gaussian Splatting.

- Główny pomysł to wykorzystanie ciągłych map dwuwymiarowych 2D, które umożliwiają efektywne podpróbkowanie w celu kontroli poziomu szczegółowości.
- Zaproponowana została progresywna strategia uczenia, która generuje mapy 2D wprost z szumu, a następnie stosuje sortowanie przestrzenne w celu uzyskania optymalnej struktury. Ponadto, w celu uzyskania wysokiej jakości renderowania, zaproponowany został specjalny modulator sterowany założonym poziomem szczegółowości renderowanych scen, który poprawia jakość wynikającą ze zmniejszonego pod-próbkowania map 2D.
- Metoda umożliwia dynamiczne skalowanie konfiguracji w zależności od mocy obliczeniowych dostępnej platformy komputerowej.

Praca ta została wykonana przez Pana mgra inż. Kacpra Kanię przy współpracy z zespołem naukowym. Doktorant jest pierwszym z jej autorów, a praca została przedłożona na konferencję ICCV 2025 (planowana na październik 2025).

Kania, K., Khirodkar, R., Saito, S., Yi, K. M., and Martinez, J., "*CLoG: Leveraging UV Space for Continuous Levels of Detail*", International Conference on Computer Vision, 2025.

Wszystkie wymienione zagadnienia badawcze, jak i ich oryginalne rozwiązania naukowe zaproponowane przez Pana mgra inż. Kacpra Kanię świadczą o Jego dużej wiedzy i dojrzałości naukowej. Świadczą one również o znajomości najnowszych osiągnięć i technologii światowych.

6. Poprawność naukowa

Czy stwierdzenia zawarte w rozprawie są godne zaufania? Czy uzasadnienia są poprawne? Wskaż zauważone słabości i błędy. Wskaż także te aspekty dotyczące poprawności, które są najbardziej wartościowe.

Pan mgr inż. Kacper Kania podjął uzasadnioną naukowo metodykę prowadzenia badań polegającą na formułowaniu problemów, analizie istniejących metod i algorytmów, a następnie na proponowaniu nowych, lepszych rozwiązań. Kolejnym etapem w jego postępowaniu było stawianie tez naukowych, które w końcu podlegały weryfikacji eksperymentalnej. Poprzez wnikliwą obserwację wyników, p. mgr inż. Kacper Kania mógł zweryfikować swoje pomysły naukowe, jak również zaproponować nowe rozwiązania. Dzięki takiemu podejściu, wszystkie zaproponowane rozwiązania, jak i tezy postawione w rozprawie, zostały zweryfikowane eksperymentalnie.

Poprawność zaproponowanych rozwiązań jest też potwierdzona poprzez publikacje naukowe z udziałem Pana mgra inż. Kacpra Kani, w *recenzowanych* periodykach o zasięgu światowym, które zostały już wymienione w poprzednim rozdziale recenzji.

Nie są to jednak jedyne publikacje Pana mgra inż. Kacpra Kani – oprócz wymienionych jest on jeszcze autorem 6 publikacji współautorskich o tematyce dotyczącej przetwarzania obrazów.

Baza publikacji Web of Science zawiera 8 publikacji, w których p. mgr inż. Kacper Kania jest jednym z autorów. Jego współczynnik tzw. współczynnik Hirscha H=3 (stan: maj 2025), co jest wynikiem ponadprzeciętnym w przypadku doktorantów.

Tym niemniej praca zawiera pewne miejsca i opisy, które wymagają dodatkowego wyjaśnienia, czy też rozszerzenia i dyskusji:

1. W pracy wiele akapitów zaczyna się od słów „*We introduce our concept ...*" (np. Rozdział 1). Jasne jest, że Doktorant pracował w zespole. Jednakże w doktoracie oceniamy indywidualny wkład właśnie Doktoranta. Dlatego też, mimo krótkiej wzmianki na temat roli i udziału współautorów w publikacjach, warto byłoby jakby Doktorant w punktach wymienił swój indywidualny wkład w każdą publikację.
2. Rozdział 2.2.2 – zaproponowana sieć neuronowa dla systemu CoNeRF to stosunkowo prosta w swojej strukturze sieć typu MLP, co jest w pewnym sensie powieleniem pomysłu z oryginalnych publikacji dotyczących metody NeRF. Ciekawe jednak czy nie lepsza byłaby tu sieć grafowa? Tego typu podejście umożliwiłoby uwzględnienie przestrzennej struktury pikseli już w samej architekturze sieci. Czy Doktorant rozważał użycie architektur sieci innych niż MLP? To samo pytanie można też powtórzyć w przypadku metody BlendFields.
3. Rozdział 2.3.3, wzór (19) – czy użyty tu symbol mnożenia to mnożenie w sensie Khatri-Rao?
4. Rozdział 2.3.4 – skoro w sieci kodujemy również pozycje punktów („*windowed positional encoding*") to nie lepiej byłoby rozważyć architekturę transformera wizyjnego, który już w naturalny dla siebie sposób dokonuje tego typu kodowania?
5. Rozdział 2.4.1 – co to znaczy „*we set the attributes to be synchronized for the training split, and desynchronized for the test split*"? Jak dokonuje się tego typu „synchronizacji" atrybutów danych?
6. Przy prezentowaniu metod warto byłoby pokazać przypadki, dla których one jednak nie działają lub działają poniżej oczekiwań. Czy jest wiele takich przypadków? Kiedy i dla jakich danych?
7. Rozdział 2.4.4 – „*our controllability losses do not interfere with the rendering quality*" – jak zostało to zmierzone?
8. W przypadku prezentowanych metod warto byłoby żeby Doktorant umieścił linki do przykładowych sekwencji ukazujących ich właściwości, np. w postaci wideo. Kilka znalazłem, ale przez przypadek.
9. Rozdział 3.2.4, rozdział 4.4.2 i wiele innych miejsc – dziwne użycie słowa „*concurrent*". Być może chodziło o „*competitive*"?
10. Rozdział 5.3.5 – nie do końca jasne jest uzasadnienie naukowe takiego, a nie innego wyboru architektury sieci MLP.
11. Obecnie obserwujemy bardzo dynamiczny rozwój badań oraz metod dotyczących połączenia, czy też sterowania procesami generatywnymi za pomocą wielkich modeli językowych LLM. Co prawda w końcowym rozdziale 6.2 Autor coś wspomina o rozwiązaniach typu „*text-prompted*", jednakże bardzo brakuje mi powiązania tego nurtu z tym co Doktorant prezentuje w swojej rozprawie.

Powyższa lista pytań dotyczy dodatkowych zagadnień naukowych, czy też sugeruje głębsze potraktowanie pewnych tematów poruszonych w rozprawie doktorskiej Pana mgra inż. Kacpra Kani. Jednakże mają one wyłącznie charakter dyskusyjny i w żadnym stopniu nie umniejszają istotnego wkładu merytorycznego przedstawionego w tej rozprawie.

7. Wiedza Kandydata

Które z rozdziałów (lub sekcji w rozdziałach) rozprawy omawiają istniejący stan wiedzy i dzięki temu potwierdzają ogólny stan wiedzy kandydata w zakresie ITT? Jakie obszary tych dyscyplin zostały omówione w tych rozdziałach/sekcjach? Jaka jest opinia recenzenta o jakości tych rozdziałów sekcji?

Jaka jest opinia recenzenta o bibliografii? Na ile bibliografia jest kompletna? Prosimy o podanie innych argumentów za lub przeciw stwierdzeniu, że kandydat posiada ogólną wiedzę w dyscyplinie informatyka techniczna i telekomunikacja.

Analizę wiedzy Doktoranta należy zacząć od analizy przedłożonej rozprawy doktorskiej. Praca ta jest dość obszerna i liczy 123 strony. Została napisana po angielsku i składa się z 6 rozdziałów oraz bibliografii. Rozpoczyna się wstępem (rozdział 1 „*Introduction*"), po czym następują 4 rozdziały będące wiernymi kopiami publikacji naukowych, które wymieniam w punkcie 5 recenzji. Ze względu na posiadane publikacje naukowe, oparcie doktoratu właśnie na nich jest zrozumiałe. Nie rozumiem jednak po co Autor żmudnie przepisywał te teksty zamiast po prostu wkleić reprinty wspomnianych artykułów, tym bardziej że porównując zarówno rozdziały, jak i ich oryginały w postaci artykułów nie zauważyłem praktycznie żadnych istotnych dodatków czy też różnic naukowych (a może szkoda?).

Pracę kończy rozdział 6 zawierający konkluzje oraz zarysowujący dalsze kierunki badań. Niemniej istotna jest bibliografia zawierająca 225 pozycji literaturowych dotyczących istotnych publikacji związanych z zagadnieniami naukowymi opisanymi w rozprawie. Jest to bardzo cenne zestawienie prac naukowych umożliwiające zarówno zagłębienie się w nowoczesne metody grafiki komputerowej, jak również zarysowujące rozwiązania uznawane już za „klasyczne" w tej dziedzinie.

Podsumowując, opracowanie nowych metod do tak ambitnego zadania wymagało od Pana mgra inż. Kacpra Kani zarówno dogłębnej znajomości już istniejących metod światowych, wiedzy zarówno w aspekcie teoretycznym, jak również w kontekście programowania oraz przetwarzania dużych zbiorów danych i organizacji eksperymentów.

8. Podsumowanie

Postawione przez Pana magistra inżyniera Kacpra Kanię tezy naukowe zostały w pełni wykazane. Opracowane przez Niego metody są nowatorskie i na bardzo wysokim poziomie naukowym. Stanowią one oryginalny własny wkład Doktoranta Pana magistra inżyniera Kacpra Kani w rozwój dyscypliny naukowej informatyka techniczna i telekomunikacja. Przy ich realizacji Pan magister inżynier Kacper Kania wykazał się dogłębną znajomością najnowszych metod i algorytmów grafiki komputerowej oraz sztucznej inteligencji, jak również dobrym opanowaniem warsztatu badawczego, a w rezultacie również wysoką dojrzałością naukową.

Recenzowaną pracę oceniam jako z spełniającą ze znacznym nadmiarem formalne oraz zwyczajowe wymagania stawiane rozprawom doktorskim. Wnioskuję o jej przyjęcie oraz o dopuszczenie Pana magistra inżyniera Kacpra Kani do publicznej obrony.

Zważywszy również na bardzo wysoki poziom naukowy opracowanych nowatorskich metod, jak również na prezentacje i publikacje tych rozwiązań podczas prestiżowych konferencji naukowych, **wnioskuję o wyróżnienie rozprawy**.

Bogusław Cyganek

Rada Naukowa Dyscypliny
INFORMATYKA TECHNICZNA
I TELEKOMUNIKACJA
Sekretariat
Data wpływu 13.06.25r.
Numer.....

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 powiązanymi 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 przygotuje 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 czworościanów 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 nauczenie 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.

Odnośnie 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 kontrolujących 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 potwierdzona. **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

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Visual Computing

ETH Zürich
CNB G105
Universitätstrasse 6
8092 Zürich
Switzerland

Prof. Dr. Marc Pollefeys
+41 44 632 31 05
+41 44 632 17 39
marc.pollefeys@inf.ethz.ch
www.inf.ethz.ch/personal/pomarc/

Zurich, 11 July 2025

PhD Thesis report:

Kacper Kania – Controllability of radiance fields with sparse information

The dissertation of Kacper Kania is focused on being able to control different aspects of neural radiance fields (NeRFs) and Gaussian splatting (GS) representations. Specifically, it aims to control facial expressions for NeRF representations of faces. Further it also aims to control lighting and level of detail for Gaussian splatting representations. The thesis consists of six chapters, including four technical chapters containing the thesis' research contributions.

The first chapter contains the introduction to the dissertation, including motivation and challenges. Further is also briefly describes the research objectives and outlines the contributions.

Chapter two introduces controllable neural radiance fields (CoNeRF). To enable control the method relies on some sparse annotations that are then propagated to the rest of the image data. It principally allows to decouple different expression changes that are seen jointly in images such that a rich controllable model can be obtained from much less data than would otherwise be required. This work was published at the IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR) in 2022, which is a top conference in the field.

Chapter three continues along the same lines of creating 3D representations of faces that can render novel expressions, but now focuses more on expression-dependent details such as wrinkles. These details can potentially be smaller than the resolution of the tetrahedra used to represent the face model. The proposed method BlendFields builds on Flame and VolTeMorph, but is shown to outperform prior art. This work was published at CVPR 2023.

The fourth chapter focuses on a different problem and switches from NeRFs to Gaussian Splatting. This more traditional graphics representation borrowed the volumetric rendering and gradient descent-based optimization approach from NeRF, but allows for much more efficient real-time rendering making it more popular. However, Gaussian Splatting, like NeRFs, normally don't handle variations in illumination in the images used to fit the appearance model. The contribution of this chapter, LumiGauss, aims to enable reconstruction from photo collections captured in the wild with varying lighting. For this purpose Kacper proposes to use 2D Gaussian Splatting to enable per view reconstruction, including surface reconstruction to enable decoupling of

appearance and lighting. The method includes the ability to handle shadows. This work was published at WACV 2025, which is a good quality conference more focused on applications of computer vision.

The last technical chapter focuses on enabling level of details for Gaussian Splatting. For this purpose the GS representation is encoded in a multi-channel 2D image where Gaussians with similar properties are clustered such that downsampling this image can generate a meaningful lower resolution representation. A decoder is then used to transform the image back into a GS representation with the desired level of detail. This method is shown to be quite effective compared to other methods. While delivering similar quality, it allows to much more effectively transition between levels of details.

The sixth chapter contains the conclusion. It discusses the impact of the work and future work.

The dissertation of Kacper Kania contains multiple solid contributions to the state of the art, several of which have been published in top computer vision conferences. Overall it is well written and provides a clear description of the research contributions. In addition, during his PhD Kacper also contributes to other publications that were not included in the thesis. All in all Kacper has delivered a solid dissertation that in my opinion definitely satisfies the requirements to obtain a doctorate from the Warsaw University of Technology.

Sincerely,