

Edycja obiektu 3D reprezentowanego przez Gaussian Splatting

W ostatnich latach obserwuje się duży wzrost oprogramowania służącego do generowania treści przez użytkowników, szczególnie w domenie wizualnej. Po premierze modelu GPT-4o od OpenAI, w ciągu zaledwie jednego tygodnia użytkownicy stworzyli ponad 700 milionów obrazów. Użytkownicy nie mających dotąd styczności z grafiką komputerową mogli w prosty sposób wygenerować lub przerobić pewien obraz według własnego upodobania. Podkreśla to ogromne zapotrzebowanie na narzędzia generatywne umożliwiające intuicyjną i bezpośrednią manipulację treściami wizualnymi.

Chociaż generowanie i edytowanie obrazów 2D staje się szybko powszechna praktyka, edycja w wyższych wymiarach, takich jak wideo, 3D i treści 4D, jest znacznie bardziej złożona. Te obszary stawiają wyzwania w zakresie spójności, geometrii, ciągłości czasowej i kontroli użytkownika, którym obecne systemy nie są jeszcze w pełni sprostać.

Gaussian Splatting (GS) to istotny postęp w grafice komputerowej, reprezentujący sceny 3D jako zbiory składników Gaussa z kolorem i przezroczystością. Jego trening i renderowanie są bardzo wydajne i generują realistyczne obrazy. GS został również zaadaptowany do dynamicznych scen 4D, obrazów 2D oraz wideo.

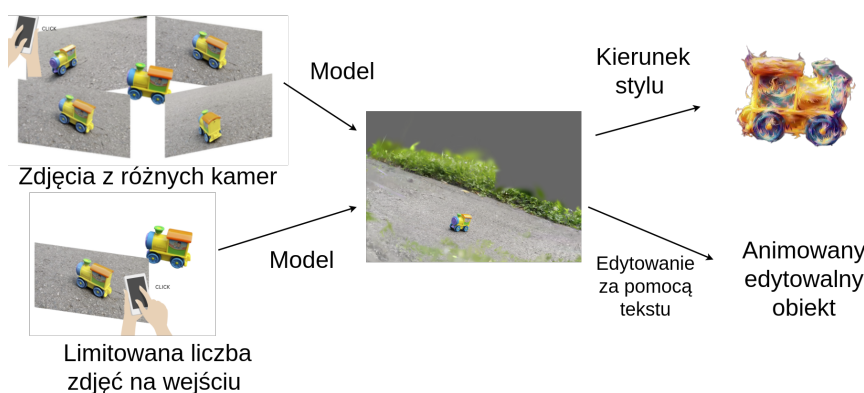


Figure 1: Przykład 3D: przy użyciu dostępnych obrazów obiekt jest rekonstruowany za pomocą modeli opartych na rozplaszczaniu gaussowskim. Użytkownik może następnie zmodyfikować właściwości wybranego obiektu.

Zareprezentowanie sceny/obiektu za pomocą prymitywów pozwala na szybki rendering, dając ogromny potencjał w zastosowaniach np. w grach 3D lub filmach. Niemniej jednak modele GS w celu rekonstrukcji obiektu/sceny potrzebują wiele zdjęć na wejściu by wiernie oddać spójność rekonstrukcji. Potencjalnie istnieje możliwość wygenerowania za pomocą dużych modeli generatywnych dodatkowych widoków na podstawie jednego zdjęcia lub tekstu. Umożliwia to dodatkowa augmentacja danych. Możliwe jest jednak wystąpienie niespójności. Problem ten potencjalnie można rozwiązać za pomocą generowania sekwencji a nie pojedynczych widoków, wykorzystując modele do generowania wideo, pozwoliłoby to dokładniejszą rekonstrukcję geometrii sceny.

Edycja obiektów reprezentowanych przez trójwymiarowe prymitywy Gaussa mieści się w czterech kategoriach: teksturyzacja, modelowanie pozycji, edycja lokalna oraz transfer stylu. Edycja kształtu i pozycji jednego prymitywu nie jest trudnym zadaniem, trudność jest w edycji w spójny sposób pewnego zbioru Gaussów z którego złożony jest obiekt. Istnieją metody, pozwalające na kontrolowanie Gaussa, niestety nie obsługują one siatki obiektu ani struktury, która pozwala na precyzyjną kontrolę ruchu obiektów 3D. Stylizacja i edycja takich scen jest jeszcze mniej trywialna, gdy poprzez ograniczone wejście modelu mamy dostęp tylko do ograniczonej liczby zdjęć pozwalających na w pełni poprawną rekonstrukcję. Potencjalnie stylizacja obiektu wpływająca na cechy prymitywów w przypadku nie spójnej rekonstrukcji mogłaby poprawić wizualny aspekt obiektu.

Projekt obejmowałby badanie możliwości rekonstrukcji obiektu 3D z pojedynczego obrazu referencyjnego lub opisu tekstowego. **Jako efekt końcowym badań przewidujemy stworzenie efektywnych i dostępnych rozwiązań umożliwiających tworzenie i edycje realistycznych modeli 3D z ograniczonym danymi wejściowymi.**