

# SWIZ

## Rekonstrukcja budynków 3D

Jakub Karpiński

Krzysztof Miśków

# Wstęp

Celem projektu jest stworzenie algorytmu rekonstrukcji 3D budynków na podstawie zdjęć z wielu kamer. Do zrealizowania projektu wykorzystany został algorytm Structure From Motion (SFM) polegający na otrzymywaniu sceny 3D na podstawie zestawu zdjęć 2D.

Algorytm SFM składa się z następujących kroków:

## 1. Ekstrakcja cech

Z każdej klatki zdjęciowej wyodrębniane są charakterystyczne punkty (cechy), takie jak narożniki budynków, krawędzie czy tekstury, za pomocą algorytmu SIFT (Scale-Invariant Feature Transform).

## 2. Dopasowanie cech

Ekstrahowane cechy są porównywane pomiędzy parami zdjęć w celu znalezienia odpowiadających sobie punktów.

## 3. Oszacowanie ruchu

Na podstawie dopasowanych punktów obliczane są ruchy kamery (jej pozycje i orientacje) w przestrzeni.

## 4. Rekonstrukcja trójwymiarowa

Gdy pozycje kamery są znane, można obliczyć trójwymiarowe położenie punktów cech za pomocą triangulacji. Każdy punkt cechy, który jest widoczny na co najmniej dwóch zdjęciach, może być zrekonstruowany w przestrzeni 3D.

## 5. Optymalizacja

W celu poprawy dokładności rekonstrukcji stosuje się technikę optymalizacji wiązki (bundle adjustment), która minimalizuje błąd reprojektacji, czyli różnicę między przewidywanymi a rzeczywistymi położeniami punktów cech na zdjęciach.

## 6. Generowanie modelu

Na podstawie punktów 3D, można wygenerować gęstą siatkę (mesh) lub chmurę punktów, która reprezentuje trójwymiarową strukturę sceny.

# Dane

Do rekonstrukcji wybrane zostały następujące datasety:

- [Fountain – P11](#)

Zawiera 11 zdjęć wykonanych z różnej perspektywy przedstawiających fontannę.

Dodatkowo dataset zawiera macierz kalibracji dla wykorzystanej kamery.



- **Castle – P30**

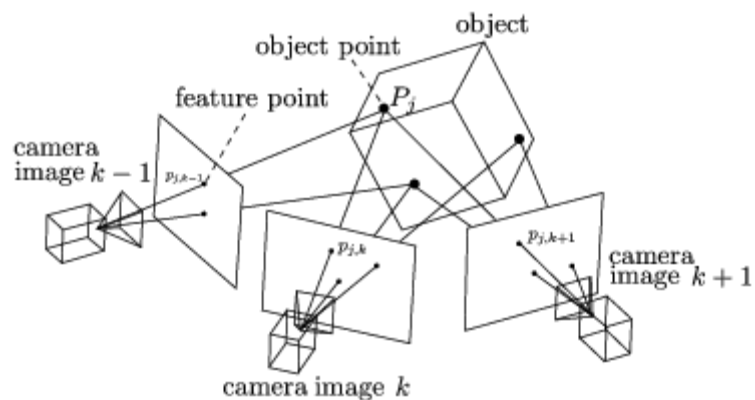
Zawiera 30 zdjęć wykonanych z różnej perspektywy przedstawiających dziedziniec zamku. Dodatkowo dataset zawiera macierz kalibracji dla wykorzystanej kamery.



## Realizacja

Podstawową ideą w rekonstrukcji obrazów 3D jest to, że mając zestaw obrazów z których każdy jest wykonany z innego punktu widzenia, możemy dokonać rekonstrukcji trójwymiarowej sceny.

Znając ruch kamer względem globalnego układu możemy określić ich ruch czyli macierze projekcji.



## Asocjacja danych

Aby to wykonać, skonstruujemy pipeline składający się z dwóch głównych części: asocjacji danych i struktury z ruchu (SfM).

Asocjacja danych jest używana do sprawdzenia, czy para obrazów jest podobna do siebie. Dwa obrazy mogą być sprawdzone pod kątem podobieństwa, używając korespondencji obrazów (SIFT) oraz geometrii dwuobrazowej (two-view geometry).

Mając zestaw nieustrukturyzowanych obrazów, najpierw znajdujemy powiązane komponenty w tych obrazach. Pomaga to znaleźć nakładające się widoki w obrazach. Aby ustanowić powiązane komponenty, używamy algorytmu **SIFT**, który pomaga nam wyodrębnić punkty kluczowe z obrazów. Następnie wykonujemy korespondencje obrazów lub punktów kluczowych, używając algorytmu geometrii dwuobrazowej.

Dzięki czemu otrzymujemy mapowanie cechy w jednym obrazie na podobną cechę w innym obrazie. Jednym z problemów w asocjacji danych jest to, że jeśli zestaw wejściowych obrazów  $N$  jest duży, to przeszukiwanie par obrazów staje się niewykonalne, a złożoność zapytania dla jednego obrazu jest rzędu  $O(N \cdot K^2)$ , gdzie  $K$  to liczba punktów kluczowych w każdym obrazie. Używana jest efektywna metoda oparta na drzewach do wyszukiwania obrazów. To sprawia, że złożoność zapytania dla obrazu redukuje się do  $O(K \cdot b \cdot L)$ , gdzie  $K$  to cechy w zapytaniu obrazu,  $b$  to gałęzie w drzewie, a  $L$  to poziomy w drzewie.

## Structure From Motion

SFM jest odpowiedzialna za początkową rekonstrukcję, używając estymacji pozycji i technik triangulacji, a następnie udoskonalenie tego przy użyciu algorytmu dopasowania wiązki. MVS jest następnie stosowany, aby uzyskać gęstą reprezentację 3D.

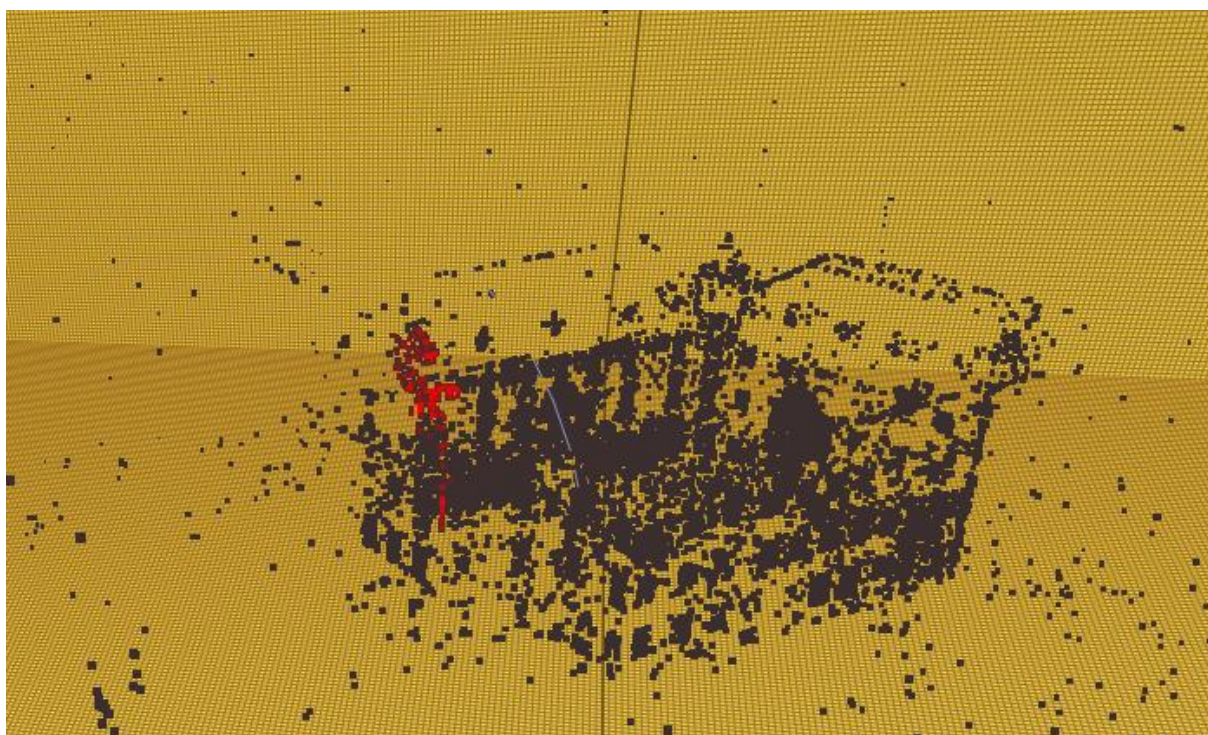
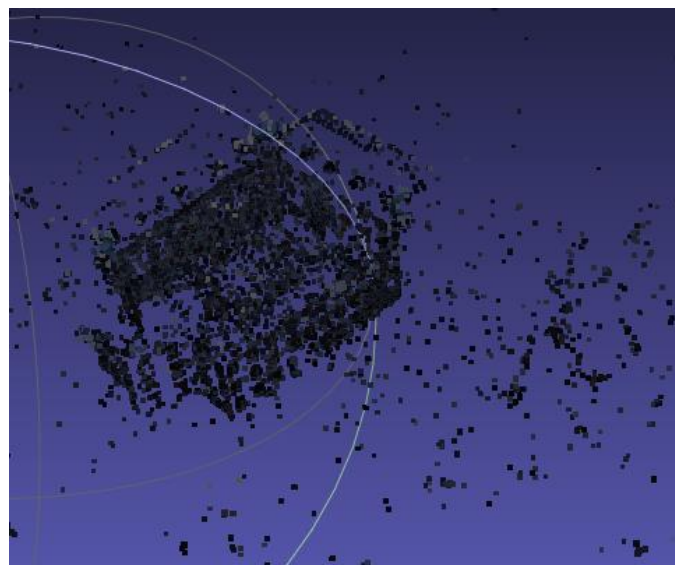
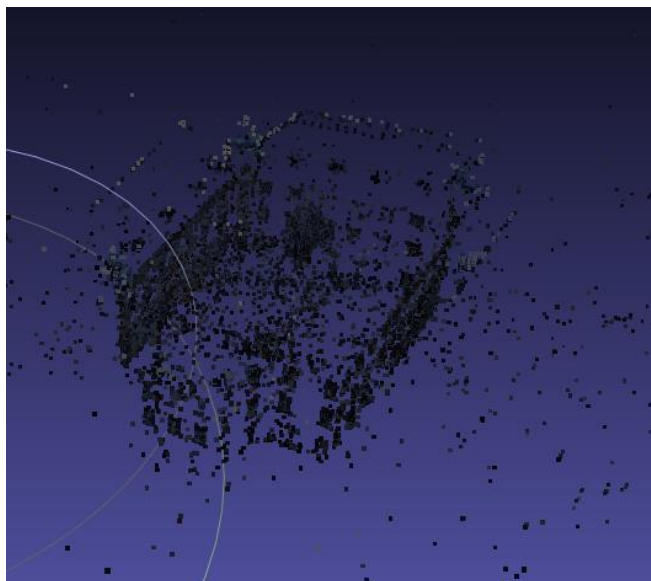
W naszym zastosowaniu użyliśmy **Incremental SFM**, która działa w następujący sposób:

- Najpierw wybieramy dwa niepanoramyczne widoki z grafu sceny wygenerowanego przez krok asocjacji danych.
- Algorytm 8-punktowy jest używany do obliczenia macierzy fundamentalnej lub istotnej. Macierz fundamentalna może być również uważana za projekcję kamery, którą można rozłożyć na dwie macierze  $P$  i  $P'$ .  $P'$  reprezentuje wewnętrzną kalibrację kamery.
- Następnie stosujemy algorytm liniowej triangulacji, aby obliczyć korespondencje i uzyskać punkty 3D.
- Algorytm dopasowania wiązki jest następnie stosowany do udoskonalenia punktów 3D uzyskanych z poprzedniego kroku.
- Następnie znajdujemy korespondencje 2D-3D i dodajemy więcej widoków do systemu. Korespondencja 2D-2D jest najpierw ustanawiana między nowo dodanym obrazem a poprzednim obrazem, a następnie ustanawia się korespondencję 2D-3D, aby uzyskać punkty 3D.
- Gdy ustanowimy korespondencje 2D-3D dla wszystkich obrazów, używamy algorytmu Perspective-n-Point (PnP), aby obliczyć pozycję obrazów względem światowych współrzędnych.
- Dodatkowo, więcej korespondencji 2D-2D może być wybranych w sąsiednich obrazach i zastosować liniową triangulację, aby znaleźć ich punkty 3D. To pomaga nam uzyskać bardziej gęstą chmurę punktów do rekonstrukcji 3D. Po uzyskaniu punktów 3D
- Następnym krokiem jest udoskonalenie tych punktów przy użyciu algorytmu dopasowania wiązki. Algorytm dopasowania wiązki jest stosowany zarówno do punktów 3D, jak i szacunków pozycji kamery uzyskanych z algorytmu PnP.



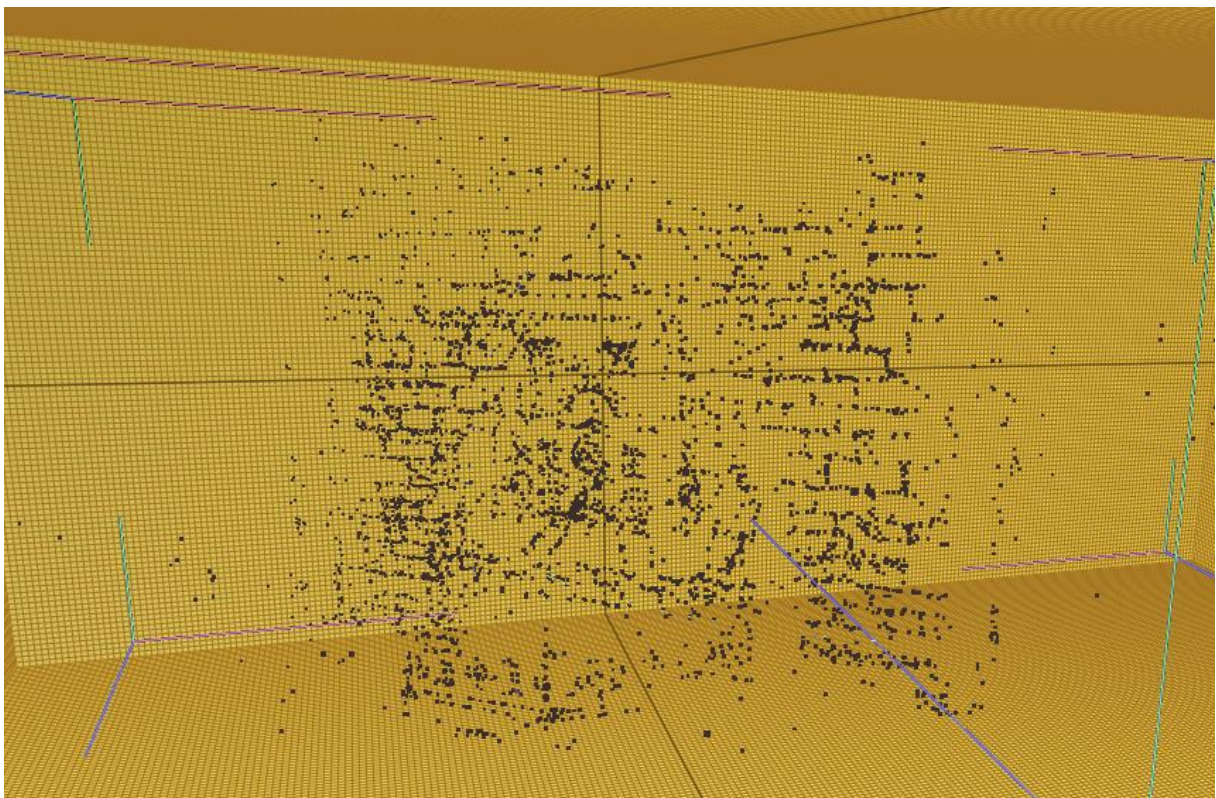
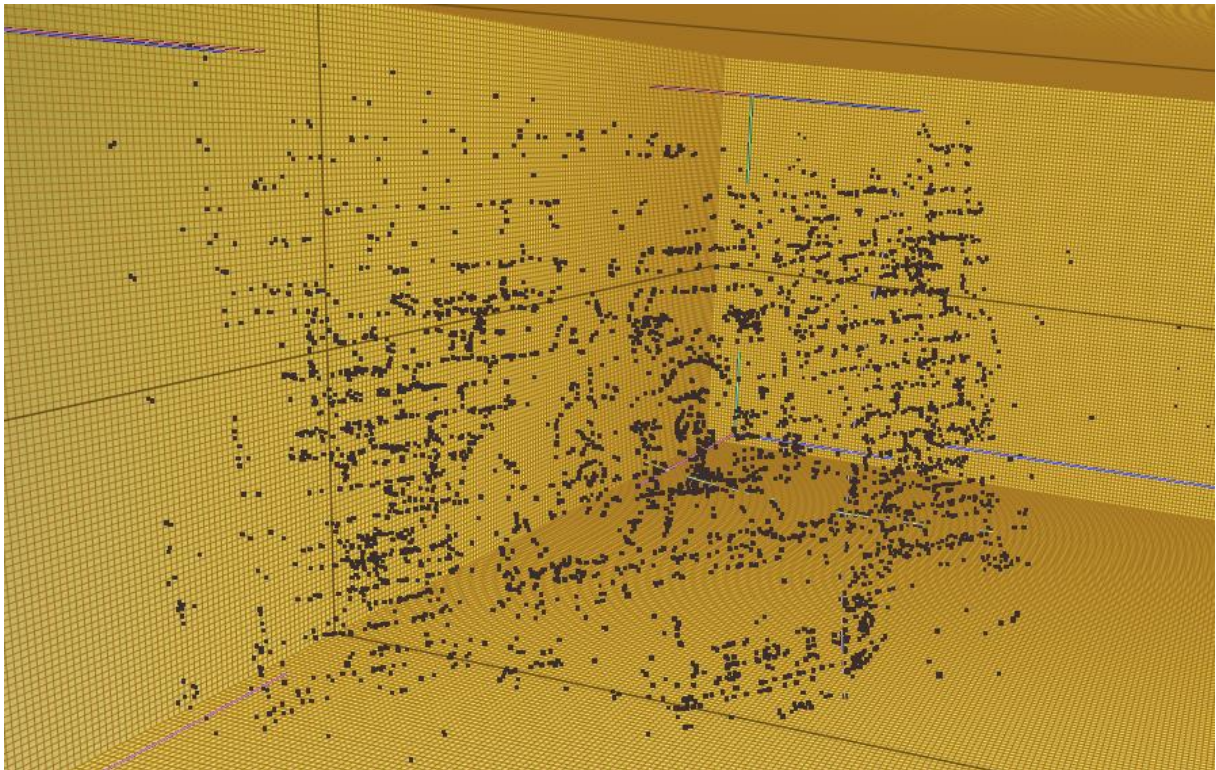
# Rezultaty

Castle- P30





## Fountain – P11



## Bibliografia

- <https://medium.com/@loboateresa/understanding-structure-from-motion-algorithms-fc034875fd0c>
- [https://www.cs.cmu.edu/~16385/s17/Slides/12.4\\_8Point\\_Algorithm.pdf](https://www.cs.cmu.edu/~16385/s17/Slides/12.4_8Point_Algorithm.pdf)
- <https://medium.com/@rashik.shrestha/perspective-n-point-pnp-f2c7dd4ef1ed>
- [https://docs.opencv.org/4.x/d5/d1f/calib3d\\_solvePnP.html](https://docs.opencv.org/4.x/d5/d1f/calib3d_solvePnP.html)
- <https://www.epfl.ch/labs/cvlab/~strecha/multiview/knownInternalsMVS.html>
- Del Pizzo, Silvio & Troisi, Salvatore. (2011). Automatic orientation of image sequences in cultural heritage. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XXXVIII-5/W16. -. 10.5194/isprsarchives-XXXVIII-5-W16-293-2011.