

POLITECHNIKA WARSZAWSKA

INŻYNIERIA MECHANICZNA
NAUKI INŻYNIERYJNO-TECHNICZNE

Rozprawa doktorska

mgr inż. Piotr Foryś

**Metoda automatycznej rekonstrukcji trójwymiarowej geometrii
obiektu dziedzictwa kulturowego na podstawie wielu pomiarów
kierunkowych.**

Promotor
prof. dr hab. inż. Robert Sitnik

Promotor pomocniczy
dr inż. Jakub Markiewicz

WARSZAWA 2023

Streszczenie

Dygitalizacja, czyli stworzenie cyfrowych reprezentacji obiektów świata rzeczywistego, znajduje zastosowanie w różnych gałęziach przemysłu, w tym medycynie, infrastrukturze krytycznej, przemyśle wojskowym, inżynierii, architekturze, rzeczywistości wirtualnej i dziedzictwie kulturowym. Nieodzownym elementem takiego procesu są skanery 3D. Proces cyfryzacji obiektów często wymaga wielu skanów z różnych pozycji, aby uzyskać rekonstrukcję pełnego, rzeczywistego modelu. W wyniku tak przeprowadzonego procesu otrzymuje się wiele kierunkowych chmur punktów, które wymagają dokładnej orientacji.

W niniejszej rozprawie przedstawiono metodę łączenia chmur punktów FAMFR (Fast Adaptive Multimodal Feature Registration). Zaproponowana koncepcja opiera się na dwóch cechach, które są wyznaczane dla każdego punktu w chmurach punktów. Pierwsza cecha opisuje kształt oraz geometrię powierzchni obiektu, a druga gradient intensywności. Obie cechy wykorzystywane są do wyznaczania punktów charakterystycznych oraz obliczania histogramów na podstawie lokalnego otoczenia. Poprzez analizę zgodności przestrzennej oraz porównywanie histogramów wybierane są najlepsze korespondencje oraz wyznaczane jest wstępne dopasowanie. Ostateczna transformacja jest otrzymywana za pomocą algorytmu Iterative Closest Point (ICP).

Dane wykorzystane w niniejszej pracy pochodzą z Muzeum Pałacu Króla Jana III w Wilanowie. Opisano także aspekt digitalizacji obiektów dziedzictwa kulturowego za pomocą skanowania 3D. Odgrywa on kluczową rolę w zachowaniu i badaniu tych obiektów. Ułatwia ekspertom ocenę stanu zabytków i jest nieodzownym wsparciem w projektach rewitalizacyjnych i konserwacyjnych w dziedzictwie kulturowym.

W pracy zestawiono oraz opisano istniejące rozwiązania dopasowywania chmur punktów. Poza klasycznymi metodami uwzględniono również metody głębokiego uczenia, które w ostatnich czasach zyskały na popularności oraz udowodniły swoją skuteczność. Walidacja metody FAMFR została przeprowadzona poprzez porównanie efektów działania z wybranymi metodami łączenia chmur punktów oraz dokładnym, manualnym dopasowaniem.

Słowa kluczowe: rekonstrukcja 3D, chmura punktów 3D, orientacja chmur punktów, skanowanie 3D, dziedzictwo kulturowe.

Abstract

Digitalization, or the creation of digital representations of real-world objects, can be found in a variety of industries, including medicine, critical infrastructure, military, engineering, architecture, virtual reality, and cultural heritage. A crucial element of this process are 3D scanners. The digitization process often requires multiple scans from different positions to obtain a complete model of the real object. As a result of this process, multiple directional point clouds are obtained, which require accurate and precise alignment (point cloud registration).

This work presents point cloud registration method called FAMFR (Fast Adaptive Multimodal Feature Registration). The proposed concept is based on two features that are estimated for each point in the point clouds. The first feature is based on geometry and the second is based on intensity gradients. Both features are used to determine characteristic points and calculate histograms based on the local environment. By analyzing the spatial correspondence and comparing the histograms, the best correspondences are selected and the initial match is determined. The precise transformation is estimated using the Iterative Closest Point (ICP) algorithm.

The data used in this work come from the Museum of King Jan III's Palace at Wilanów. Additionally, the aspect of digitization of cultural heritage objects using 3D scanning was also described. It plays a key role in preserving and studying these objects. This process allows experts to assess the condition of cultural heritage object and is an indispensable support in revitalization and conservation projects in cultural heritage.

The paper summarizes and describes the state-of-the-art solutions for point clouds registration process. In addition to classic methods, deep learning methods have also been included, which have recently gained popularity and proven their effectiveness. Validation of the FAMFR method was carried out by comparing the results with selected methods of point cloud registration and accurate, manual adjustment.

Keywords: 3D reconstruction, 3D point cloud, point cloud registration, 3D scanning, cultural heritage.

Tabela skrótów i pojęć

Skrót	Pełna forma
TLS	Terrestrial Laser Scanning
LiDAR	Light Detection and Ranging
UAV	Unmanned Aerial Vehicle
DK	Dziedzictwo Kulturowe
HBIM	Historical Information Building Modelling
SHM	Structural Health Monitoring
ICP	Iterative Closest Point
RANSAC	Random Sample Consensus
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SI	Spin Images
PPF	Point Pair Feature
PFH	Point Feature Histogram
FPFH	Fast Point Feature Histogram
PPFRGB	Point Pair Feature Red Green Blue
PFHRGB	Point Feature Histogram Red Green Blue
VFH	Viewpoint Feature Histogram
SHOT	Signature of Histograms of Orientations
3DSC	3D Shape Context
LRF	Local Reference Frame
USC	Unique Shape Context
CORS	Concentric Ring Signature
MI	Moment Invariants
NBS	Normal Based Signature
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
PCE	Principal Curvature Estimation
HWR	Haar Wavelet Response
PIG	Point Intensity Gradient

Skrót	Pełna forma
HOG	Histogram of Oriented Gradients
RIFT	Rotation Invariant Feature Transform
ISE	Intensity Spin Estimation
SLAM	Simultaneous Localization and Mapping
VSLAM	Visual Simultaneous Localization and Mapping
SLAM6D	Six-Dimensional SLAM
RGBD	Red Green Blue Depth
DNN	Deep Neural Network
CGF	Compact Geometric Features
PPFNet	Point Pair Feature Network
AE	Auto Encoder
GPS	Global Positioning System
INS	Inertial Navigation System
LK	Lucas and Kanade algorithm
SVD	Singular Value Decomposition
DCP	Deep Closest Point
DGCNN	Dynamic Graph Convolutional Neural Network
PRNet	Partial Registration Network
PCRNet	Point Cloud Registration Network
FCGF	Fully Convolutional Geometric Features
DGR	Deep Global Registration
LMVD	Learning multiview 3D point cloud registration
DSAC	Differentiable RANSAC
USIP	Unsupervised Stable Interest Point
FPN	Feature Proposal Network
Predator	Pairwise point cloud registration with deep attention to the overlap region
CGF	Compact Geometric Features
DIP	Distinctive 3D Local Deep Descriptors
YOHO	You Only Hypothesize Once
IMFNet	Interpretable Multimodal Fusion Network

Skrót	Pełna forma
DAM	Descriptor Activation Mapping
PointDSC	Deep Spatial Consistency
SM	Spectral Matching
CSCE-Net	Channel-Spatial Contextual Enhancement Network
TEASER	Truncated least squares Estimation And SEMidefinite Relaxation
TLS	Truncated Least Squares
RCP	Rectified Corresponding Points
DOF	Degrees of Freedom
SC2-PCR	Second order spatial Compatibility for Point Cloud Registration
RPM-Net	Robust Point Matching Network
RGM	Robust Graph Matching
GNN	Graph Neural Network
PCAM	Product of Cross-Attention Matrices
DeepGMR	Deep Gaussian Mixture Registration
IDAM	Iterative Distance-Aware Similarity Matrix Convolution
OMNet	Overlapping Mask Network
GeoTransformer	Geometric Transformer
PCFD	Progressive and Coherent Feature Drift
UTOPIC	UncerTainty-aware Overlap PredICtion
CPD	Coherent Point Drift
GMM	Gaussian Mixture Model
EM	Expectation-Maximization
UME	Universal Manifold Embedding
TL	Transfer Learning
SDV	Smoothed Density Value
R-PCR	Recurrent Point Cloud Registration
GRU	Gate Recurrent Unit
GCNet	Geometry-guided Consistent Network
FMR	Feature Matching Recall
RR	Registration Recall
RMSE	Root Mean Square Estimation

Skrót	Pełna forma
FAMFR	Fast Adaptive Multimodal Feature Registration
FRAMES	Framework and Robust Algorithms for Models of Extreme Size
LoD	Level of Detail
NLP	Natural Language Processing
NCBiR	Narodowe Centrum Badań i Rozwoju
SL	Structured Light
SfS	Shape from Silhouette
SfM	Structure from Motion
CPU	Central Processing Unit

Tabela 1: Tabela skrótów i pojęć.

Spis treści

1	Wstęp	11
1.1	Motywacja	11
1.2	Cel i założenia pracy	15
1.3	Układ pracy	17
2	Przegląd istniejących rozwiązań	18
2.1	Metody łączenia chmur punktów	18
2.1.1	Metody oparte na cechach charakterystycznych	19
2.1.2	Metody iteracyjne	26
2.1.3	Metody oparte na głębokim uczeniu	28
2.2	Podsumowanie	44
3	Charakterystyka danych wejściowych	49
3.1	Opis skanowanych pomieszczeń zabytkowych	49
3.2	System pomiarowy	53
3.3	Dane	54
4	Koncepcja	61
5	Implementacja	64
5.1	Środowisko implementacyjne	64
5.2	Przetwarzanie wstępne	68
5.3	Wyznaczanie punktów charakterystycznych	70
5.4	Histogramy cech	72
5.5	Wyznaczanie korespondencji	73
5.6	Parametry	75
6	Ocena skuteczności metody FAMFR	81
6.1	Kryterium oceny	81
6.2	Eksperyment	83
6.3	Wyniki	84
7	Dyskusja	88

7.1	Wyzwania oraz ograniczenia metody FAMFR	88
7.2	Analiza wyników	91
8	Podsumowanie	95
8.1	Realizacja celów pracy	96
8.1.1	Realizacja celów naukowych	96
8.1.2	Realizacja celu aplikacyjnego	97
8.2	Potencjalne dodatkowe zastosowania	100
8.3	Przyszłe prace	102
9	Bibliografia	103

1 Wstęp

Obecny trend wskazujący na coraz większe zapotrzebowanie na digitalizację obiektów świata realnego, na ich reprezentację w postaci cyfrowych bliźniaków (ang. digital twins) spowodował bardzo duży i szybki wzrost dostępności systemów do skanowania trójwymiarowego (3D). Systemy te wykorzystują metody pomiarowe takie jak: naziemne skanery TLS (Terrestrial Laser Scanning) [1], LiDAR (Light Detection and Ranging) [2], fotogrametria bliskiego zasięgu [3] czy światło strukturalne [4]. W idealnym przypadku, system wykorzystywany do przeprowadzania procesu cyfryzacji powinien umożliwiać precyzyjne wymiarowanie geometrii trójwymiarowych obiektów wraz z informacjami o ich barwie. Często, wynikiem takiego skanu, jest chmura punktów. Chmura punktów to zbiór punktów w trójwymiarowej przestrzeni, które reprezentują powierzchnię obiektu. Każdy punkt w chmurze posiada trójwymiarowe współrzędne (x , y , z) oraz często dodatkowe atrybuty takie jak kolor (np. w postaci wartości RGB), wektory normalne, a w przypadku skanerów TLS intensywność odbicia wiązki lasera. Innymi słowy, chmura punktów jest efektem cyfryzacji geometrii 3D realnego obiektu, który w pełni odzwierciedla jego wielkość oraz kształt. Pojedyncza chmura może składać się z wielu milionów punktów. Dzięki chmurom punktów 3D możliwe jest tworzenie precyzyjnych rekonstrukcji modeli obiektów, które znacząco ułatwiają analizę ich kształtu, struktury oraz przeprowadzenie wirtualnych pomiarów. Systemy pomiarowe oraz wynikowe chmury punktów znajdują zastosowania w bardzo różnych gałęziach przemysłu, takich jak: medycyna [5], infrastruktura krytyczna [6], przemysł wojskowy [7], architektura [8], rozrywka wirtualna [9] oraz dziedzictwo kulturowe [10].

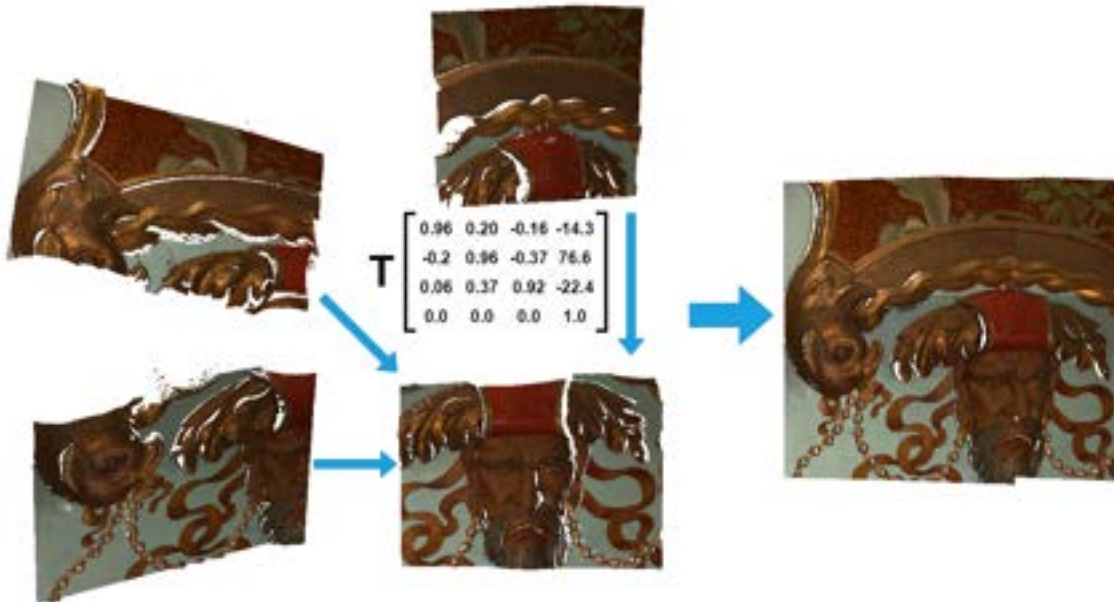
1.1 Motywacja

Zapotrzebowanie na wysokiej jakości wizualizację i dokumentację obiektów dziedzictwa kulturowego w znaczący sposób sprzyja rozwojowi systemów skanowania 3D, takich jak: ultralekkie, bezzałogowe statki powietrzne (UAVs, ultra-light Unmanned Aerial Vehicles [11]), urządzenia takie jak skanery laserowe czy triangulacyjne [12], smartfony z wbudowanymi sensorami LiDAR [13] czy oprogramowanie do fotogrametrii [14]. Precyzyjne skany pozwalają dokładnie udokumentować stan obiektów dziedzictwa kulturowego, umożliwiając badanie ich kondycji oraz analizę [15], a zaawansowane narzędzia znacząco ułatwiają tworzenie dokładnych modeli cyfrowych. Rozwój tych systemów

spowodował zwiększenie liczby projektów dokumentacyjnych, konserwacyjnych, czy prac związanych z procesem rewitalizacji zabytków oraz miejsc dziedzictwa kulturowego. Coraz więcej osób może angażować się w proces cyfrowego dokumentowania i zachowywania dziedzictwa kulturowego, co prowadzi do większego zrozumienia i docenienia naszej wspólnej historii i kultury. Wygenerowana w ten sposób dokumentacja odgrywa kluczową rolę w ułatwianiu badań nad obiektami i miejscami dziedzictwa kulturowego, przyczyniając się istotnie do postępów w tej dziedzinie. Udokumentowane modele stanowią istotne narzędzie dla ekspertów z różnych regionów świata, wspierając ich w ocenie i analizie stanu zachowania zabytków.

Ponadto, systematycznie powtarzane procesy skanowania 3D w czasie pozwalają śledzić zmiany w obiektach, które mogą być wynikiem wpływu różnych czynników, takich jak warunki atmosferyczne, środowisko czy działalność człowieka [16, 17]. Ważnym aspektem jest również ochrona, utrzymanie i konserwacja miejsc dziedzictwa kulturowego, a do tego niezbędna jest właśnie odpowiednia dokumentacja [18]. Technologia skanowania 3D dostarcza szczegółowe dane, które znacząco ułatwiają dalsze badania naukowe, planowanie działań konserwatorskich oraz, w razie potrzeby, rekonstrukcję obiektów. Ponadto, skany te mają ogromne znaczenie edukacyjne, umożliwiając społeczeństwu korzystanie z wirtualnych wizualizacji i rekonstrukcji oraz ułatwiają dzielenie się wiedzą na temat dziedzictwa kulturowego [19, 20, 21]. Jednym z rezultatów skanowania 3D jest możliwość generowania chmury punktów. Wysokiej jakości chmury punktów 3D są często niezbędne do dokładnego utrwalenia niezwykle, skomplikowanych szczegółów obiektów dziedzictwa kulturowego [22, 23, 24]. Chmury punktów stały się nieodzownym elementem digitalizacji i nowoczesnej dokumentacji obiektów dziedzictwa kulturowego (DK). Są one wykorzystywane do zadań związanych z modelowaniem (HBIM - Historical Building Information Modelling), monitorowaniem (SHM - Structural Health Monitoring), detekcją uszkodzeń czy wirtualną rekonstrukcją [25]. Pozwalają one dokładnie odwzorować złożoną geometrię, barwę oraz teksturę powierzchni obiektu. Chmury punktów znajdują zastosowanie w wielu dziedzinach, takich jak: klasyfikacja [26], segmentacja [27], modelowanie [28], czy omawiana dokumentacja 3D [29]. W praktyce dziedzictwa kulturowego często konieczne jest łączenie wielu chmur punktów w celu uzyskania spójnego i pełnego modelu 3D. Dane z jednej pozycji skanera często są niewystarczające, aby odwzorować cały obiekt, dlatego często wykorzystuje się algorytmy do orientacji chmur punktów (ang. Point Cloud Registration). Jest to proces,

który polega na dopasowaniu i połączeniu dwóch lub więcej chmur punktów za pomocą odpowiedniej transformacji 3D, tak aby umieścić je we wspólnym układzie odniesienia, którym może być lokalny układ odniesienia związany z opracowywanym obiektem lub globalny układ odniesienia związany z Państwowym Układem Współrzędnych Geodezyjnych (Rysunek 1).



Rysunek 1: Wizualizacja procesu łączenia chmur punktów w jeden obiekt.

Poza prostym przypadkiem, gdzie występują jedynie dwie chmury punktów wymagających połączenia, są projekty wymagające bardziej złożonego procesu do łączenia wielu chmur punktów zebranych z różnych kierunków [30]. Wybór odpowiedniej metody zależy od wielkości obiektu, projektu oraz liczby chmur do przetworzenia, a także możliwości rozmieszczenia sygnalizowanych punktów odniesienia na obiekcie oraz w jego najbliższym otoczeniu. Metody dopasowywania wielu chmur punktów, poza znajomością parametrów orientacji chmury punktów, wymagają dodatkowo określenia kolejności ich przetwarzania. Jednakże nawet niewielkie błędy dopasowania między dwoma sąsiednimi chmurami punktów mogą prowadzić do kumulacji i propagacji błędów. Dlatego jakość wyznaczania względnych orientacji pomiędzy chmurami punktów jest bardzo ważna, gdyż wpływa na jakość ostatecznego modelu. Finalny model jest również poddawany całościowemu wyrównaniu (ang. Bundle Adjustment) w celu wyznaczenia finalnych elementów orientacji chmur punktów, co prowadzi do bardziej dokładnego i spójnego przedstawienia obiektu.

Przykładowy proces łączenia wielu chmur punktów zebranych z różnych perspektyw widokowych znajduje się na rysunku 2.

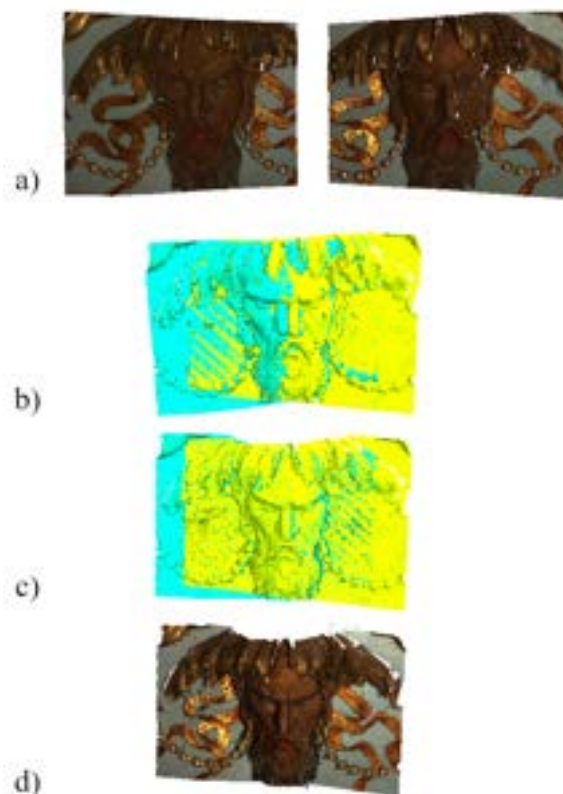


Rysunek 2: Wizualizacja procesu dopasowywania wielu chmur punktów z różnych perspektyw.

Wyzwanie związane z łączeniem chmur punktów DK często wymaga zastosowania zaawansowanych algorytmów, znacznej ilości czasu i wiedzy eksperckiej. Obiekty dziedzictwa kulturowego różnią się między sobą pod względem wykonania, tekstury, kształtu i stanu zachowania. Obiekty te powstawały w różnych okresach historycznych i były przechowywane w różnych warunkach. Dodatkowo, skanowanie 3D tych obiektów może być przeprowadzane przy użyciu różnych urządzeń i metod, które generują różnorodne rodzaje danych, charakteryzujące się zróżnicowaną gęstością, szumem pomiarowym i dokładnością. Zmienne warunki oświetleniowe oraz częściowo odblaskowe i przezroczyste powierzchnie mogą prowadzić do niedoskonałości w odwzorowaniu kolorów i geometrii, znacząco utrudniając proces łączenia chmur punktów [31]. Aby sprostać tym wyzwaniom, konieczne jest dostosowanie metod do konkretnych wymagań, uwzględniając sposób pozyskania danych oraz specyfikę danego obiektu dziedzictwa kulturowego.

Aby uzyskać jak najbardziej dokładne i kompleksowe modele, należy dokonać odpowiedniego przeglądu dostępnych technik przetwarzania danych wraz z doбором odpowiedniego algorytmu łączenia chmur punktów [32]. Powstało wiele metod i strategii łączenia danych 3D w postaci chmur punktów. Są to na przykład metody łączące chmury

punktów TLS z wykorzystaniem różnych odmian algorytmu ICP (Iterative Closest Point [33]), wykorzystujące geometryczne cechy powierzchni do łączenia chmur punktów pochodzących z dużych i szczegółowych obiektów dziedzictwa kulturowego [34], metody automatycznego dopasowania wykorzystujące informację pochodzące z obrazów 2D [35] czy metody wykorzystujące wiele istniejących cech opisujących geometrię lub kolor chmury punktów [36]. Często taki proces składa się z dwóch etapów: pierwszy z nich obejmuje znalezienie wstępnego dopasowania chmur punktów, a drugi wykorzystuje metody iteracyjne, do końcowego dopasowania dwóch chmur punktów, rysunek 3.



Rysunek 3: Proces orientacji chmur punktów; a) dane wejściowe, b) wstępne dopasowanie, c) precyzyjne łączenie d) finalny model.

1.2 Cel i założenia pracy

Celem postawionym w niniejszej rozprawie jest opracowanie skutecznej metody łączenia chmur punktów, będących efektem skanowania pomieszczeń obiektów dziedzictwa kulturowego, wraz z implementacją algorytmów i ich weryfikacją. Zadanie to zostało podzielone na dwie grupy celów do zrealizowania:

Cele naukowe

1. Opracowanie skutecznej metody łączenia pary chmur punktów bazującej na cechach geometrycznych oraz intensywności wyznaczonej z koloru o parametrach uzależnionych od średniej odległości między punktami D_{avg} :
 - kontrolny błąd dopasowania $RMSE < 1.5 \cdot D_{avg}[mm]$ - błąd średniokwadratowy szacowany na podstawie odległości między punktami kontrolnymi na transformowanej i referencyjnej chmurze punktów,
 - różnica wyliczonego dystansu Chamfera [37] pomiędzy wynikiem (D_{CH}), a wartością referencyjną (D'_{CH}): $D'_{CH} - D_{CH} < 3.0 \cdot D_{avg}[mm]$ - Odległość ta jest obliczana poprzez zsumowanie kwadratów odległości między najbliższymi odpowiadającymi sobie punktami w dwóch chmurach punktów,
 - różnica współczynnika podobieństwa S (wzór 8) pomiędzy wynikiem (S_V), a wartością referencyjną (S'_V): $S'_{Vg} - S_{Vg} < 0.15$ & $S'_{Vs} - S_{Vs} < 0.15$ (15% maksymalnej wartości dla cechy kształtu S_{Vs} i gradientów S_{Vg}).
2. Opracowany algorytm musi być zdolny do poprawnego działania z różnymi typami danych wejściowych, takimi jak chmury punktów reprezentujące płaskie ściany z bogatymi zdobieniami i malowidłami, płaskorzeźby oraz obiekty o złożonym kształcie, przy uwzględnieniu obecności licznych złożonych i błyszczących dekoracji.

Cel aplikacyjny

Celem aplikacyjnym jest skuteczne zastosowanie tej metody do opracowania efektywnego i szybkiego procesu do łączenia dużej ilości wysoko-rozdzielczych chmur punktów, które są efektem kampanii skanowania 3D, które odbyło się w Muzeum Pałacu Króla Jana III w Wilanowie. W wyniku prac digitalizacyjnych zostały wytworzone kierunkowe chmury punktów 3D dla następujących pomieszczeń:

- Antygabinet Królowej i Gabinet al Fresco - dokumentacja w ramach projektu „Restauracja i zabezpieczenie symbolu polskiego dziedzictwa kulturowego - Muzeum Pałacu Króla Jana III w Wilanowie” współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Infrastruktura i Środowisko,

- Gabinet Chiński Króla (pomiar po zakończeniu prac konserwatorskich) i Garderoba Króla – dokumentacja w ramach projektu „Rewitalizacja i digitalizacja jedynej w Polsce barokowej rezydencji królewskiej w Wilanowie” (Program Operacyjny Infrastruktura i Środowisko) finansowany ze środków Unii Europejskiej i Ministerstwa Kultury i Dziedzictwa Narodowego.

Aby zrealizować ten cel, opracowana metoda powinna spełnić następujące wymagania:

- skrócenie czasu potrzebnego na stworzenie pełnego modelu o co najmniej 50% - przy założeniu, że pełny model składa się z 50 chmur punktów, czas liczony w stosunku do manualnego łączenia danych z wykorzystaniem ogólnodostępnego algorytmu ICP. Wybór 50 chmur punktów jest uwarunkowany przez automatyczny proces skanowania pomieszczenia, który był wykonywany fragmentami. Pojedynczy fragment składa się z 50 chmur punktów. Drugim argumentem jest czas potrzebny do przeprowadzenia eksperymentu.
- automatyzacja i parametryzacja algorytmów dopasowania poprzez uzależnienie ich parametrów sterujących od charakterystyki wejściowych chmur punktów.

1.3 Układ pracy

Rozprawę doktorską podzielono na dziewięć rozdziałów. Pierwszy rozdział jest wprowadzeniem, opisuje motywację, cel oraz założenia pracy, a także w skrócie przedstawiono w nim układ pracy. Rozdział drugi zawiera przegląd istniejących metod automatycznej orientacji chmur punktów: metody klasyczne, oparte na szukaniu punktów charakterystycznych oraz obliczaniu deskryptorów, metody iteracyjne oraz nowe metody oparte na głębokim uczeniu. Rozdział trzeci opisuje dane wejściowe wykorzystane w pracy, skąd pochodzą, jak zostały pozyskane i czym się charakteryzują. Czwarty rozdział przybliża koncepcję zaproponowanej metody, a rozdział piąty opisuje jej implementację wraz z wykorzystanymi narzędziami. Rozdział szósty przedstawia kryterium oceny wraz z eksperymentem oraz wynikami walidującymi metodę FAMFR. W rozdziale siódmym przeprowadzono dyskusję dotyczącą analizy wyników oraz wyzwań i ograniczeń zaproponowanej metody. Rozdział ósmy podsumowuje całą rozprawę, skupiając się na realizacji postawionych celów oraz potencjalnych dodatkowych zastosowań wraz z możliwymi kierunkami rozwoju. Rozdział dziewiąty zawiera spis bibliograficzny.

2 Przegląd istniejących rozwiązań

Wybór właściwej metody ma kluczowe znaczenie dla łączenia chmur punktów. Metody oparte na geometrii mogą być odpowiednie tylko w przypadku skanów posiadających kształt i szczegółową geometrię powierzchni, a chmury punktów przedstawiające płaskie fragmenty, z bogatą teksturą wymagają analizy informacji o teksturze/kolorze w celu precyzyjnego dopasowania. W większości przypadków mamy jednak do czynienia z połączeniem obu tych cech w pojedynczej chmurze punktów, dlatego należy wziąć pod uwagę kolor i kształt obiektu. Nakładające się regiony i części wspólne skanów 3D również mogą stanowić wyzwanie, zwłaszcza gdy wzajemne pokrycie między przetwarzanymi chmurami punktów jest zbyt małe, aby zastosowana metoda mogła go skutecznie przetworzyć. Dodatkowo, metody można podzielić ze względu na charakter procesu dopasowywania: istnieją metody wstępnie łączące chmury punktów oraz metody finalnego dopasowania, które wymagają, aby chmury punktów posiadały już wstępną estymację transformacji.

2.1 Metody łączenia chmur punktów

Na przestrzeni lat opracowano różnorodne metody łączenia chmur punktów. Techniki łączenia chmur punktów można podzielić na trzy główne kategorie: metody klasyczne, oparte na szukaniu i dopasowywaniu punktów kluczowych na podstawie deskryptorów (znane również jako hand-crafted) [38], metody iteracyjne, takie jak metoda ICP (Iterative Closest Point [39, 40]) wraz z jej modyfikacjami oraz metody oparte na sieciach neuronowych i głębokim uczeniu. Każdą kategorię można sklasyfikować na pomniejsze grupy (Rysunek 4).

Metody klasyczne są dobrze znane i mają szerokie zastosowanie, często stanowią też punkt wyjścia dla nowszych rozwiązań opartych na głębokim uczeniu. Metoda ICP, będąca przedstawicielem metod iteracyjnych, jest powszechnie stosowana w łączeniu chmur punktów, rekonstrukcji trójwymiarowego środowiska, lokalizacji robotów i planowaniu trasy. Metoda ta, do efektywnego działania, wymaga wstępnego dopasowania. Jej główną wadą jest czas obliczeń, który może być zbyt długi dla niektórych zastosowań. Obecnie wiele nowych rozwiązań korzysta z sieci neuronowych i głębokiego uczenia [41]. Choć są one bardzo szybkie, wymagają dużej ilości danych uczących. Są one skuteczne w wielu zastosowaniach przemysłowych [42, 43], gdzie obiekty mają powtarzalną geometrię i teksturę, jednak nie są one odpowiednie do łączenia chmur punktów dziedzictwa kulturowego. Obiekty te często są

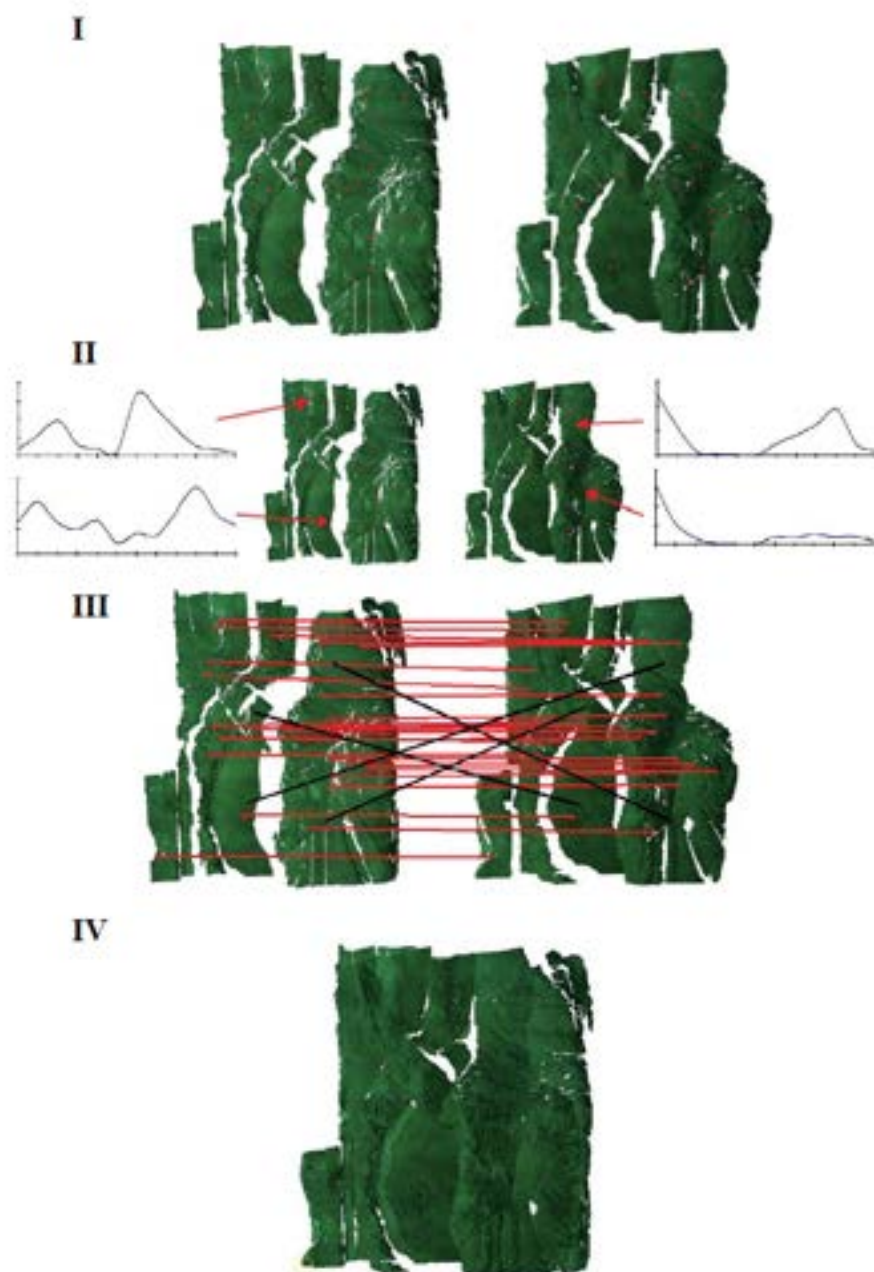


Rysunek 4: Metody łączenia chmur punktów.

unikalne, posiadają zróżnicowaną teksturę, bogate zdobienia i wyszukaną geometrię, co utrudnia tworzenie odpowiednich zbiorów uczących. Dodatkowo wymagają one często długotrwałego procesu uczenia, który nie zawsze daje satysfakcjonujące wyniki, wymaga douczania i jest wrażliwy na przetrenowanie. Ponadto, nie istnieje powszechnie dostępna baza danych zawierająca wystarczającą ilość chmur punktów obiektów dziedzictwa kulturowego gotowych do wytrenowania uniwersalnego modelu opartego na sieciach neuronowych.

2.1.1 Metody oparte na cechach charakterystycznych

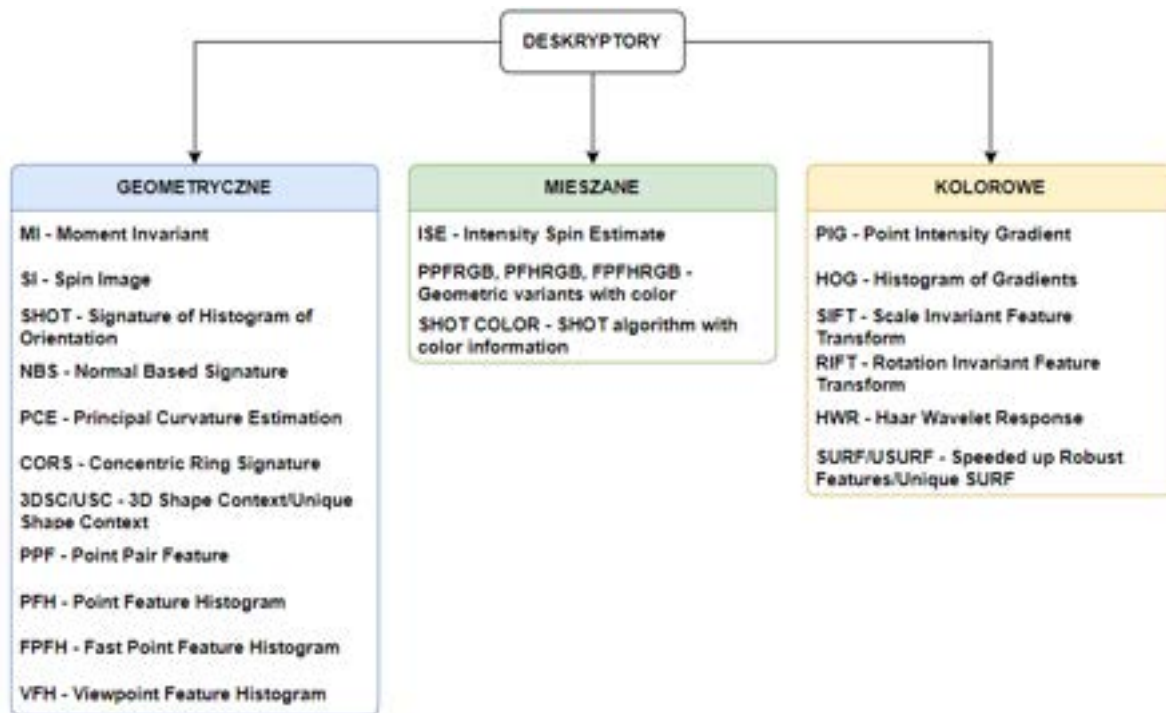
Połączenie dwóch chmur punktów jest zadaniem składającym się z kilku części, takich jak szukanie punktów charakterystycznych, wyznaczenie wartości deskryptorów, porównanie par punktów i wyszukiwanie korespondencji, wyznaczenie elementów orientacji wzajemnej tj. trzech kątów obrotu i trzech parametrów transformacji liniowej chmur punktów oraz ocena wyników (Rysunek 5).



Rysunek 5: Etapy łączenia chmur: I - wyznaczenie punktów charakterystycznych, II - wyliczenie deskryptorów punktów charakterystycznych, III - wyszukiwanie najlepszych połączeń, IV - połączenie dwóch chmur punktów.

Każdy z wymienionych etapów wymaga wykorzystania odpowiedniego algorytmu. Proces dopasowywania chmur punktów wymaga znalezienia oraz opisu punktów kluczowych na wszystkich chmurach, które chcemy połączyć. Algorytmy różnią się skutecznością, czasem wykonywania oraz dokładnością. Deskryptory, czyli algorytmy opisujące chmurę punktów, opierają się głównie na lokalnym otoczeniu, ale istnieją również deskryptory globalne. Mogą

one wykorzystywać najróżniejsze cechy chmury punktów takie jak geometrię, powierzchnię, kolor, intensywność lub korzystać z kilku cech jednocześnie (rysunek 6).



Rysunek 6: Podział deskryptorów ze względu na cechy.

Tak wyznaczone oraz opisane punkty charakterystyczne są porównywane w celu znalezienia najlepszych dopasowań. Odpowiadające sobie pary punktów tworzą tzw. korespondencje. Następnie wykorzystywane są algorytmy odporne (ang. robust) bazujące na analizie zależności geometrycznych, pozwalające na wykrycie obserwacji odstających spowodowanych błędnym dopasowaniem deskryptorów. Przykładem takiego algorytmu jest RANSAC (Random Sample Consensus [44]). Z wynikowego zbioru poprawnych dopasowań estymowana jest finalna transformacja.

Punkty charakterystyczne

Na przestrzeni czasu opracowano wiele algorytmów detekcji punktów charakterystycznych i deskryptorów je opisujących. Są one często używane w połączeniu jako detektor-deskryptor w celu identyfikacji unikalnego obszaru obiektu widocznego we wspólnych fragmentach danych pomiarowych pozyskanych z różnych kierunków. Dzięki temu połączeniu możliwe jest stworzenie pełnego wirtualnego obrazu przedmiotu. Punkt charakterystyczny to punkt w obrazie 2D lub 3D, który wyróżnia się pod względem geometrii, koloru lub tekstury

w porównaniu do najbliższego otoczenia. Ma on matematycznie uzasadnioną definicję i zdefiniowane położenie. Jego otoczenie zawiera bogate informacje, co umożliwia skuteczne odróżnienie go od innych punktów. Dodatkowo, pożądaną cechą punktu charakterystycznego jest duża odporność na lokalne i globalne zakłócenia oraz szum pomiarowy, co przyczynia się do wysokiego stopnia powtarzalności obliczeń. Detektory punktów charakterystycznych do wyszukiwania punktów kluczowych w chmurach punktów 3D wykorzystują geometrię i/lub informację o kolorze. Istnieje wiele algorytmów do wykrywania punktów charakterystycznych, a wśród najczęściej używanych możemy wymienić SIFT [45, 46], SURF [47, 48], ISS [49] i Harris3D [50].

Deskrytory geometryczne

SI (Spin Images [51]) jest lokalnym deskrytorem kształtu, wykorzystywanym między innymi do rozpoznawania wielu obiektów na scenach, klasyfikacji i modelowania. Tworzy on charakterystyczny obraz zdefiniowany przez dwie współrzędne cylindryczne α i β , który następnie jest obracany wokół osi określonej przez zorientowany punkt oraz skojarzony z nim wektor normalny. Punkty w lokalnym otoczeniu są rzutowane na tak utworzony obraz.

Następna grupa metod opiera się na relacji między punktami w sąsiedztwie, a ich wektorami normalnymi. Ta relacja, nazywana PPF (Point Pair Feature [52]), opisuje względną pozycję, orientację i odległość między parą punktów. Istnieje wiele wariantów tej cechy. Obejmują one PFH (Point Feature Histogram [52]), który koduje geometryczne właściwości danego punktu oraz jego sąsiedztwa poprzez uogólnienie średniej krzywizny w punkcie, wykorzystując wielowymiarowy histogram. FPFH (Fast Point Feature Histogram [38]) jest obliczany w podobny sposób, jednak zamiast używać wszystkich sąsiednich par punktów, jest obliczany tylko dla kluczowego punktu i jego sąsiadujących punktów. FPFH jest budowany z kilku uproszczonych instancji PFH, co znacząco przyspiesza czas potrzebny na jego wyznaczenie. VFH (Viewpoint Feature Histogram [53]) jest rozszerzeniem deskryptora PFH o punkt widzenia. Tak jak w przypadku PFH, geometria jest reprezentowana przez histogram, który dodatkowo zawiera komponent punktu widzenia. Histogram opisujący ten komponent jest tworzony poprzez zebranie kątów pomiędzy kierunkiem widzenia, a wszystkimi wektorami normalnymi punktów chmury obiektu. VFH jest deskrytorem globalnym, opisującym cały obiekt.

Innym deskrytorem wykorzystującym histogram do opisu cechy jest SHOT (Signature of

Histograms of Orientations [54]). Deskryptor dla danego punktu charakterystycznego jest obliczany poprzez wyznaczenie lokalnych histogramów z informacjami geometrycznymi na temat lokalizacji punktów wewnątrz siatki sferycznej z uwzględnieniem wektorów normalnych.

Deskryptor 3DSC (3D Shape Context [51]) to rozszerzona wersja standardowego deskryptora kształtu 2D na trzy wymiary. Wykorzystuje on siatkę sferyczną tworzącą układ horyzontalny skupioną wokół punktu charakterystycznego. Siatka ta jest podzielona na komórki wzdłuż wymiarów promienia oraz wartości azymutu i elewacji. Wartość pojedynczej komórki siatki jest ważoną sumą punktów znajdujących się w jej objętości. Wadą tej metody jest brak określonego lokalnego układu odniesienia (LRF, Local Reference Frame), który negatywnie odbija się na efektywności tej metody. Dlatego też zaproponowano rozszerzenie tego deskryptora, USC (Unique Shape Context [54]), który oblicza jeden lokalny układ odniesienia dla każdego punktu charakterystycznego, co jednoznacznie określa orientację siatki sferycznej.

Deskryptor CORS (Concentric Ring Signature [55]) reprezentuje lokalną topologię 3D. Konstrukcja tego deskryptora polega na wyszukiwaniu wszystkich punktów sąsiadujących, znajdujących się w lokalnej sferze, obliczaniu średniej płaszczyzny projekcji, wyznaczeniu lokalnego układu odniesienia na płaszczyźnie oraz obliczaniu komórek. Tak wyznaczony kontener może zostać przedstawiony w formie macierzy, która jest uzupełniana projekcją sąsiadujących punktów na płaszczyźnie. Wartością dodawaną do każdej komórki jest odległość między punktem, a powierzchnią interpolowaną przez cztery sąsiadujące komórki.

Deskryptor Moment Invariants (Moment Invariants [56]) wylicza 3 niezmienniki momentów w danym punkcie 3D, które są niezależne od translacji, zmian w skali oraz rotacji. Pierwszym etapem tworzenia deskryptora jest znalezienie centroidu na podstawie punktów sąsiadujących z punktem charakterystycznym, a następnie wyliczenie trzech najczęściej wykorzystywanych niezmienników momentów centralnych.

Deskryptor NBS (Normal Based Signature [57]) w danym punkcie tworzy lokalny układ odniesienia (u, v, n) , gdzie n jest wektorem normalnym skojarzonym z wybranym punktem. Następnie definiowany jest lokalny dysk wokół punktu charakterystycznego na płaszczyźnie skorelowanej z wektorem normalnym punktu. Na tak utworzony dysk projektowane są punkty, a ich odległości do dysku są wpisywane w utworzony kontener. Finalny wektor cech jest uzyskiwany przez zastosowanie DCT (dyskretna transformata cosinusowa) i DFT (dyskretna

transformata Fouriera) na stworzonym kontenerze.

PCE (Principal Curvature Estimation [58]) szacuje główne krzywizny powierzchni w wybranych kluczowych punktach, obliczając ich kierunki (wektory własne) i wielkości (wartości własne). Wynikowy wektor cech składa się z pięciu elementów, w którym trzy elementy reprezentują główny kierunek krzywizny (wektor własny powiązany z największą wartością własną) i dwie wartości: maksymalną i minimalną wartość własną. Wartości te uzyskuje się z dekompozycji macierzy kowariancji utworzonej przy użyciu wektorów normalnych chmury.

Deskryptory oparte na kolorze

Jedną z grup deskryptorów koloru są metody oparte na falkach Haara [59]. Pierwszym z nich jest HWR (Haar Wavelet Response [47, 60]). Ten deskryptor jest rozszerzeniem klasycznych falek Haara do trzech wymiarów. Metoda jest analogiczna do przypadku dwuwymiarowego, jednak w tym przypadku filtr jest sferą. Sfera jest podzielona na dwie strony: dodatnią i ujemną. Aby uzyskać dodatnią odpowiedź, sumuje się intensywność wszystkich punktów znajdujących się po jednej stronie sfery. Podobnie, ujemną odpowiedź otrzymuje się przez zsumowanie intensywności punktów po przeciwnej stronie. Ostateczny deskryptor jest sumą wszystkich odpowiedzi dla każdego punktu w trzech kierunkach (X, Y, Z) ważonych funkcją Gaussa. Kolejne dwie metody, SURF (Speeded up Robust Features) i U-SURF (Upright-SURF [47, 60]) również opierają się na falach Haara. Podobnie jak w przypadku deskryptora HWR, odpowiedzi falek Haara są obliczane w trzech kierunkach. Wszystkie odpowiedzi są przechowywane w lokalnym układzie współrzędnych, który następnie służy do wyznaczenia dominującego kierunku poprzez sumowanie wszystkich odpowiedzi znajdujących się w skanowanym otoczeniu. Kolejnym krokiem jest zbudowanie kontenera; jest to trójwymiarowy sześciąt składający się z 27 identycznych komórek (sześciątów) zorientowany w dominującym kierunku. W każdej komórce siatki znajdują się wszystkie punkty, a następnie obliczany jest sześćo-elementowy wektor cech reprezentujący dodatnie oraz ujemne odpowiedzi falek Haara. W przypadku U-SURF, nie wyznacza się dominującego kierunku, co znacznie przyspiesza działanie algorytmu.

Duża grupa deskryptorów opiera się na gradientach intensywności wyznaczanych dla punktu charakterystycznego. PIG (Point Intensity Gradient [61]), polega na estymacji gradientu intensywności w danym obszarze. Gradient w danym punkcie jest wektorem

prostopadłym do płaszczyzny kierunku normalnego, w której występuje największy wzrost lokalnej intensywności. Wielkość tego wektora wskazuje zmianę intensywności. Drugim deskryptorem opartym na gradientach jest rozwinięcie deskryptora PIG o nazwie HOG (Histogram of Oriented Gradients [61]). Cecha opisana jest histogramem kątów między znormalizowanym gradientem w punkcie charakterystycznym, a wektorem utworzonym z tego punktu i jego sąsiadującego punktu. Trzeci deskryptor oparty na gradientach, SIFT (Scale Invariant Feature Transform [61]), wykorzystuje tę samą metodę co poprzednie deskryptory do obliczenia gradientu. W pierwszym etapie następuje estymacja lokalnego układu współrzędnych. Dla każdego punktu wyznaczony jest gradient reprezentujący lokalną zmianę intensywności w trzech kierunkach. Każdy obliczony gradient jest przyporządkowywany do układu współrzędnych, a następnie estymowany jest główny kierunek gradientu. W ostatnim etapie budowany jest trójwymiarowy sześcian składający się z 27 identycznych komórek (sześciąt) ustawionych w dominującym kierunku. Lokalny histogram gradientów jest obliczany w każdej komórce siatki przy użyciu analogicznej metody, jak w przypadku deskryptora HOG. Czwarty deskryptor oparty na gradientach, RIFT (Rotation Invariant Feature Transform [62]), stanowi rozwinięcie standardowego deskryptora RIFT w kontekście trzech wymiarów. W ramach tego rozszerzenia, wokół punktu kluczowego tworzone są cztery sfery o różnych promieniach, a następnie dokonywana jest analiza histogramu orientacji gradientu dla każdej z tych sfer, zgodnie z metodologią przyjętą w deskrypcji HOG.

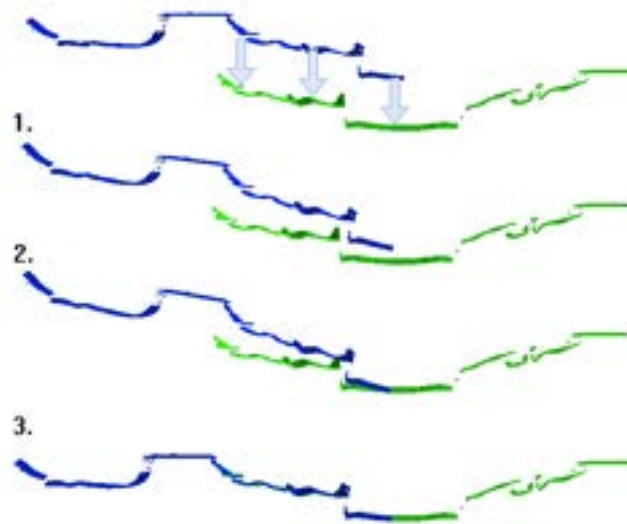
Deskryptory mieszane

Deskryptory, które opisują jednocześnie kolor i kształt chmury punktów są często rozwinięciem swoich odpowiedników, które opisują tylko jedną z tych cech. Grupa deskryptorów oparta o PPF posiada swoje warianty, które poza kształtem kodują informację o kolorze. Są to PPFRGB (Point Pair Feature RGB [63]), PFHRGB (Point Feature Histogram RGB [63]). Dodanie informacji o kolorze czyni te deskryptory bardziej przydatnymi dla chmur punktów zawierających zarówno kolor, jak i kształt. Innym przypadkiem jest algorytm SHOT, dla którego wariancja tego deskryptora (SHOTColor [54]) dodaje informacje kolorowe oparte na przestrzeni barw CIELab. ISE (Intensity Spin Estimation [64]) opiera się na deskrypcji SI, jednak zamiast kształtu wykorzystuje informację dotyczącą intensywności światła. Deskryptor ten jest reprezentowany za pomocą dwuwymiarowego histogramu, który zawiera rozkład wartości jasności w sąsiedztwie punktu kluczowego.

2.1.2 Metody iteracyjne

ICP

ICP (Iterative Closest Point [39, 40]) oraz jego warianty stanowią fundamenty łączenia danych oraz rekonstrukcji 3D. Metoda ta jest powszechnie stosowana w dziedzinach widzenia komputerowego, robotyki oraz dopasowania dwóch chmur punktów. Polega ona na iteracyjnym znalezieniu najlepszej transformacji dopasowującej jedną chmurę punktów do drugiej poprzez minimalizację odległości pomiędzy odpowiadającymi sobie punktami. Istnieją różne warianty tej metody z których najpopularniejszymi są punkt-do-punktu oraz punkt-do-płaszczyzny. Niestety, ICP jest wrażliwe na szum pomiarowy oraz wymaga wstępnego, przybliżonego dopasowania w celu poprawnego połączenia dwóch chmur punktów. Dodatkowo, może ona wpadać w lokalne minimum, co całkowicie uniemożliwia poprawne dopasowanie. Metoda jest bardzo dokładna, ale również wymagająca obliczeniowo, szczególnie dla wysoko-rozdzielczych chmur punktów. ICP dzięki swojej skuteczności znalazła bardzo szerokie zastosowanie oraz istnieje bardzo wiele najróżniejszych odmian tego algorytmu. Klasyczny przykład łączenia dwóch chmur punktów jest pokazany na rysunku 7.



Rysunek 7: Schemat łączenia dwóch chmur punktów za pomocą algorytmu ICP minimalizując odległość w 3 iteracjach; etap 3 reprezentuje finalną transformację, etapy 1-2 są krokami pośrednimi, w każdym coraz bardziej minimalizowana jest odległość pomiędzy chmurami punktów.

Warianty algorytmu ICP różnią się w zależności od danych wejściowych jak i metryki. Cechami, które są wyróżniane w algorytmie mogą być:

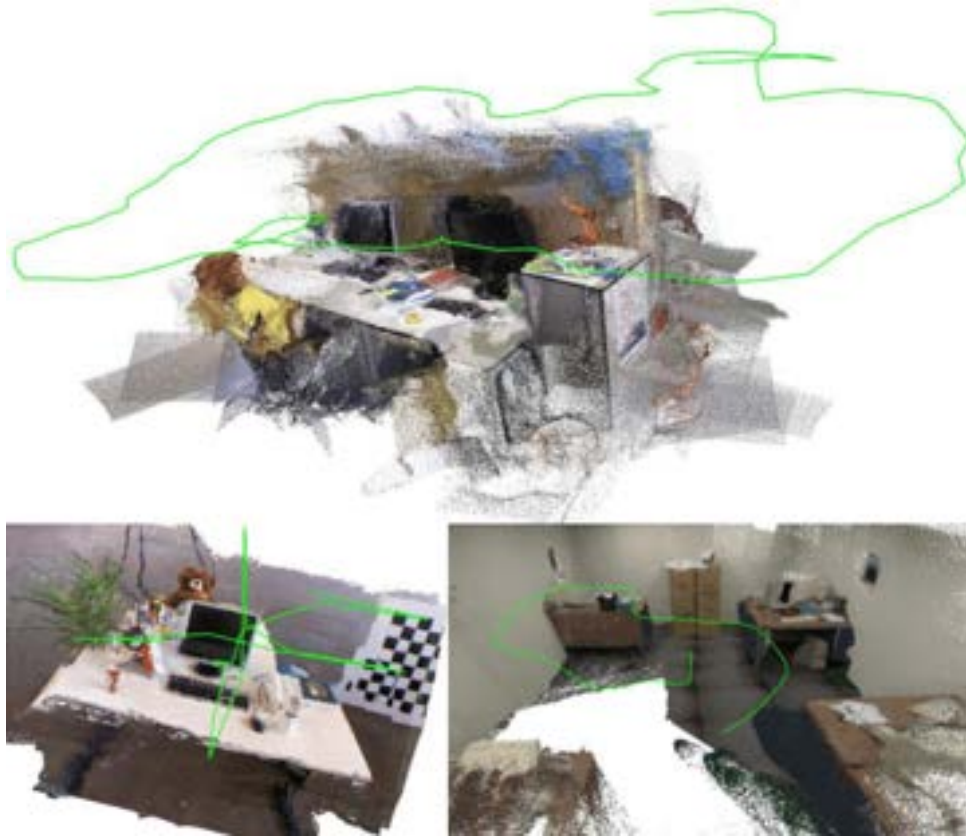
- wybór punktów - wszystkie, różne, wybrane na podstawie np. określonej metryki,
- porównanie, waga, odrzucanie obserwacji odstających (ang. outlier) - mogą być oparte o konkretne metryki; może to być wartość stała, dystans pomiędzy punktami, podobieństwo kształtu na podstawie wektorów normalnych, podobieństwo kolorów na podstawie wartości RGB, intensywności, gradientów itp.,
- metryka - punkt-do-punktu, punkt-do-płaszczyzny itp.,
- minimalizacja błędu - linowa & nie-linowa optymalizacja (np. Levenberg-Marquardt [65]).

SLAM/SLAM6D

SLAM (Simultaneous Localization and Mapping [66]) jest metodą obliczeniową szeroko stosowaną w robotyce. Jest wykorzystywana do jednoczesnego mapowania otoczenia i lokalizacji robota w zadanym środowisku. SLAM może wykorzystywać najróżniejsze sensory do pobierania danych ze środowiska, takich jak LiDAR, sensory akustyczne, kamery. SLAM w połączeniu z danymi pochodzącymi z sensorów wizyjnych nazywamy VSLAM (Visual Simultaneous Localization and Mapping). VSLAM dzięki swojej efektywności i szybkości w rekonstrukcji posiada wiele wariantów przystosowanych do różnych zastosowań. Dodatkowo wymagania sprzętowe dla tej metody są niskie, dlatego często wykorzystywana jest w systemach czasu rzeczywistego. Wykorzystanie wymienionej metody razem z dronami (UAV) pozwala na szybką i efektywną rekonstrukcję mierzonego obszaru [67]. VSLAM można znaleźć w różnego rodzaju aplikacjach, gdzie istotne jest odtworzenie trójwymiarowego modelu środowiska.

SLAM6D (Six-Dimensional SLAM [68]) jest rozszerzeniem wyżej wymienionej metody na dodatkowe wymiary często związane z ruchem agenta/robota oraz otoczeniem. Przykładowymi parametrami może być: prędkość, prędkość kątowna, przyspieszenie. Wykorzystanie tych parametrów w znaczący sposób ułatwia odwzorowanie aktualnego stanu robota oraz analizę środowiska w jakim się znajduje. Uwzględnienie dodatkowych wymiarów pozwala na wszechstronniejsze i dokładniejsze oszacowanie trajektorii oraz mapy, jednakże wiąże się to z wykorzystaniem bardziej wymagających algorytmów i czujników, co wpływa

istotnie na czas obliczeń. SLAM 6D ma zastosowanie w różnych dziedzinach, w tym w robotyce, pojazdach autonomicznych, rzeczywistości rozszerzonej i wirtualnej. Może być szczególnie przydatny w scenariuszach, w których dodatkowe wymiary odgrywają istotną rolę, jak na przykład śledzenie szybko poruszających się obiektów lub nawigacja w dynamicznym środowisku.



Rysunek 8: Przykłady rekonstrukcji wykorzystujące system SLAM [69].

2.1.3 Metody oparte na głębokim uczeniu

Choć tradycyjne metody łączenia danych 3D mają swoje zastosowania i korzyści, metody oparte na głębokim uczeniu i sieciach neuronowych oferują nowe możliwości, zwłaszcza w przypadku dużych zbiorów danych, różnorodnych źródeł informacji i złożonych scenariuszy, gdzie automatyzacja i elastyczność są kluczowe. Rozwój metod głębokiego uczenia i opartych na sieciach neuronowych pokazały interesującą, alternatywną ścieżkę w stosunku do tradycyjnych rozwiązań w łączeniu danych 3D. Ich atrakcyjność w głównej mierze opiera się na zdolności automatycznego i efektywnego ekstrakowania cech 3D z chmury punktów oraz późniejszego ich dopasowania bez potrzeby ręcznego definiowania

funkcji podobieństwa. Dzięki wykorzystaniu technik głębokiego uczenia oraz nowych kart graficznych proces ten może być stosunkowo szybki i efektywny. Inną dodatkową cechą jest możliwość integracji różnych danych wejściowych, takich jak obrazy, chmury punktów czy informacje o głębi. Pozwala to uzyskać bardziej dokładny i kompleksowy model 3D. Coraz to nowsze rozwiązania dobrze radzą sobie również w przypadku zaszumionych danych, braku informacji o fragmentach obiektu czy zmiennymi warunkami oświetleniowymi dzięki uczeniu na różnych poziomach abstrakcji oraz automatycznemu wyodrębnianiu istotnych cech pod kątem procesu łączenia danych. Skalowalność oraz adaptacyjność do zmieniających się warunków pozwalają na zastosowanie ich do dużych zbiorów danych i złożonych problemów.

Wiele nowoczesnych metod łączenia chmur punktów opartych na głębokim uczeniu koncentruje się wyłącznie na informacjach geometrycznych, z pominięciem informacji o teksturze. Istnieją jednak pewne wyjątki, gdzie te metody wykorzystują pośrednie media, takie jak obrazy RGBD, obrazy projekcyjne lub mapy głębi [70, 71, 72]. Informacje o teksturze pozwalają człowiekowi na rozróżnienie poszczególnych części sceny. Ze względu na to, że metody głębokiego uczenia zazwyczaj przetwarzają jedynie względne położenia punktów, brakuje im informacji o kolorze, co wpływa na ich ograniczone zastosowanie. W kontekście łączenia chmur punktów związanych z dziedzictwem kulturowym, obiekty te często charakteryzują się złożonymi detalami i bogatymi zdobieniami, które mogą mieć różne kolory i tekstury. Dlatego też, pomijanie informacji o kolorze w procesie dopasowywania, negatywnie wpływa na rezultaty łączenia chmur punktów, co może prowadzić do uzyskania zdecydowanie mniej dokładnych wyników. Ponadto, metody głębokiego uczenia stosowane w łączeniu chmur punktów DK posiadają pewne ograniczenia, takie jak konieczność posiadania znacznej ilości danych treningowych oraz potencjalne możliwości do przetrenowania. Obecne deskryptory punktów oparte na głębokim uczeniu często są postrzegane jako czarne skrzynki, które nie pokazują w jaki sposób oryginalne punkty są przetwarzane w celu generacji końcowego deskryptora.

Metody łączenia chmur punktów 3D opartych na głębokim uczeniu można względnie podzielić na wiele różnych kategorii. Przykładem może być podział zaproponowany w artykule przeglądowym [73], gdzie autorzy zdecydowali się pogrupować metody na dwie kategorie:

- Detection-description-matching - obejmującą wszystkie metody (klasyczne oraz oparte na głębokim uczeniu) biorące udział w procesie łączenia chmur punktów poprzez

wybranie punktów kluczowych na podstawie określonego kryterium, wyznaczenie opisu cech każdego z punktów na podstawie jego lokalnego otoczenia, wykorzystanie odpowiednich strategii do wyszukania korespondencji pomiędzy punktami, a następnie znalezienia odpowiedniej transformacji.

- All-in-one - metody, które przyjmują zupełnie inną strategię estymującą transformację. Są to rozwiązania, które zawierają w sobie wszystkie niezbędne komponenty, aby z wejścia składającego się z dwóch chmur punktów otrzymać transformację niezbędną do poprawnego dopasowania.

Do pierwszej kategorii można zakwalifikować metody, które zostały opisane w poprzednim rozdziale, a dodatkowo deskrytory oparte na głębokim uczeniu, takie jak: PointNet, 3DMatch, CGF, PPFNet, PPF-FoldNet, 3DFeat-Net. PointNet [74] jest pierwszą próbą wykorzystania sieci DNN (Deep Neural Network [75]) z chmurami punktów jako danymi wejściowymi w celu klasyfikacji oraz segmentacji. Mimo swej prostoty, architektura PointNet osiąga bardzo dobre wyniki i zrewolucjonizowała rozwój kolejnych metod sieci neuronowych wykorzystujących chmury punktów. 3DMatch [76] to model, który wykorzystuje uczenie maszynowe do nauczenia się lokalnego deskryptora geometrycznego niezbędnego do poprawnego ustalania korespondencji. Deskryptor ten jest reprezentacją cech lokalnej geometrii i umożliwia porównywanie i dopasowywanie punktów pomiędzy różnymi fragmentami danych 3D. Autorzy zaproponowali metodę uczenia się cech (supervised learning [77]), która opiera się na wykorzystaniu milionów etykiet korespondencji znalezionych w istniejących rekonstrukcjach scen RGB-D. 3DMatch korzysta z trójwymiarowej konwolucyjnej sieci neuronowej (ConvNet [78]), która przyjmuje lokalną objętość wokół dowolnego punktu zainteresowania na powierzchni 3D i oblicza deskryptor cechy dla tego punktu. Deskryptor CGF (Compact Geometric Features [79]) to deskryptor cech geometrycznych, który reprezentuje lokalną geometrię w chmurach punktów o nieuporządkowanej strukturze. W odróżnieniu od innych podejść, CGF nie wymaga uporządkowanej i spójnej parametryzacji powierzchni, reprezentacji wolumetrycznych lub syntezy dodatkowych obrazów głębi. Jest w stanie procesować dane w przestrzeni euklidesowej. PPFNet, czyli Point Pair Feature Network [80], to model uczenia głębokiego, który skupia się na ekstrakcji lokalnych cech z chmur punktów 3D w celu ustalenia korespondencji między nimi. PPFNet opiera się na wcześniej wspomnianej architekturze PointNet. W podejściu wykorzystywanym w PPFNet reprezentacja lokalnej geometrii opiera

się na kombinacji punktów oraz wektorów normalnych, zwanych cechą pary punktów (PPF). Dzięki wykorzystaniu zaproponowanej funkcji straty (N-tuple loss) do trenowania, metoda osiąga się bardzo dobre wyniki i ma możliwość wprowadzenia kontekstu globalnego do procesu uczenia. Inną metodą również wykorzystującą cechy PPF jest PPF-FoldNet [81]. W przeciwieństwie do wcześniejszego podejścia, PPF-FoldNet nie uwzględnia pierwotnych punktów ani wektorów normalnych w kodowaniu. Sieć składa się z auto-enkodera (AE), który łączy kodowanie PointNet z dekodерem FoldingNet [82]. AE jest trenowany w celu odtworzenia cech PPF za pomocą metryki odległości zbioru. Architektura PPF-FoldNet opiera się na PointNet w celu osiągnięcia niezależności od permutacji. Proces trenowania jest prostszy w porównaniu do innych metod, takich jak 3DMatch, ponieważ nie wymaga próbkowania par lub trójek z wcześniej oznaczonego zbioru danych. 3DFeat-Net [83] to częściowo nadzorowana architektura uczenia głębokiego, która integruje detekcję oraz deskrypcję cech 3D na podstawie chmur punktów 3D. Architektura 3DFeat-Net opiera się na strukturze Syjamskiej [84], która uczy się rozpoznawać, czy dwie chmury punktów 3D pochodzą z tego samego miejsca dzięki informacji pochodzącej z połączenia dwóch technologii GPS (Global Positioning System) oraz INS (Inertial Navigation System) [85]. Wykorzystuje ona również PointNet jako mechanizm wejściowy do bezpośredniego przetwarzania chmur punktów 3D.

Druga z kategorii obejmująca metody głębokiego uczenia zawiera takie metody jak: PointNetLK, DeepICP, DCP, PRNet, PCRNet. Metoda PointNetLK [86, 87] łączy algorytm Lucasa i Kanade (LK [88]) oraz model PointNet w celu stworzenia pojedynczej rekurencyjnej głębokiej sieci neuronowej. Algorytm nie wymaga wyznaczania korespondencji pomiędzy punktami, co w rezultacie daje znaczne korzyści pod względem dokładności i efektywności obliczeniowej. PointNetLK posiada zdolność do efektywnego łączenia chmur punktów nawet dla obiektów i kształtów, które nie były uwzględnione w procesie trenowania. DeepICP to architektura oparta na uczeniu maszynowym do precyzyjnego dopasowywania dwóch różnych chmur punktów z pominięciem etapu odrzucania wartości odstających (ang. outliers) oraz estymowania transformacji metodą RANSAC. Metoda wyodrębnia semantyczne cechy każdego punktu zarówno z chmury punktów źródłowej, jak i docelowej, korzystając z sieci ekstrakcji cech PointNet++ [89]. Następnie, punkty są opisywane deskryptorem cech za pomocą mini-PointNet [74] w celu zwiększenia dokładności dopasowania. Dzięki wykorzystaniu odległości euklidesowej L2 jako funkcji straty, adaptacyjne uwzględnianie wag

punktów kluczowych oraz wykorzystaniu operatora SVD (Singular Value Decomposition [90]) z biblioteki TensorFlow [91] możliwa jest jedno-iteracyjna estymacja korespondencji pomiędzy punktami. Metoda DCP (Deep Closest Point [92]) wykorzystuje sieć DGCNN (Dynamic Graph Convolutional Neural Network [93]) do nauki korespondencji oraz różniczkowalną procedurę rozkładu wartości (SVD) w celu łączenia chmur punktów 3D. Dzięki modelowi składającemu się z trzech części: sieci do kodowania chmury punktów, modułu opartego o uwagę (ang. attention-based [94]) z warstwą generacji cech oraz różniczkowalnej warstwy SVD metoda pozwala efektywnie estymować dopasowanie dwóch chmur punktów. DCP wykazuje obiecujące wyniki nawet dla chmur punktów nie uwzględnionych podczas treningu. Należy zaznaczyć jednak, że ta metoda zakłada dokładne i jednoznaczne dopasowanie pomiędzy dwoma rozkładami chmur punktów. Ten warunek nie zawsze jest spełniony podczas pracy przy rzeczywistych scenariuszach, gdzie chmury punktów mogą posiadać wysokie wartości szumu pomiarowego lub braki w rekonstrukcjach. PRNet (Partial Registration Network [95]) jest architekturą opartą o moduł detekcji punktów charakterystycznych. Jej głównymi założeniami jest wykrywanie punktów charakterystycznych, które występują w obu chmurach punktów, a następnie dopasowanie tych punktów charakterystycznych poprzez rozwiązanie problemu Prokrusta [96]. Metoda PRNet jest iteracyjna, co pozwala na stopniowe doskonalenie wstępnego dopasowania. PRNet wyróżnia się także tym, że jest metodą samo-nadzorowaną (ang. self-supervised learning [97]). Ponadto, nauczone reprezentacje mogą być wykorzystane również do zadań klasyfikacji. PCRNet (Point Cloud Registration Network [98]) wykorzystuje PointNet w architekturze Syjamskiej do kodowania informacji o kształcie chmur punktów źródłowych oraz docelowych jako wektory cech. Następnie, analizując oraz interpretując dane, estymuje się transformację, która dokładnie dopasowuje te dwa wektory cech. PCRNet wykorzystuje informacje specyficzne dla kształtu już w fazie treningowej, co czyni ją bardziej odporną na szum w danych. Dodatkowo, PCRNet korzysta z podejścia PointNetLK, które wykorzystuje klasyczne techniki dopasowywania, takie jak algorytm Lucas-Kanade (LK), do dopasowywania cech PointNet.

Podobną propozycję podziału metod zaproponowali autorzy w artykule "Deep learning based point cloud registration: an overview"[99]. Metody zostały sklasyfikowane na metody bez korespondencji oraz metody bazujące na korespondencjach (wraz z podziałem na moduły: ekstraktor cech, szukanie korespondencji, odrzucanie wartości nietypowych/odstających,

estymacja transformacji). Dodatkową kategorią zaproponowaną przez autorów jest grupa metod poprawiających efektywność dopasowania poprzez zastosowanie mechanizmów iteracyjnych lub korzystających z ograniczeń wynikających z analizowania względnych relacji pomiędzy wieloma chmurami punktów. Metody należące do tej dodatkowej kategorii nie są tematem niniejszej rozprawy, ponieważ wymagają wstępnego dopasowania. Dlatego zostały one pominięte w opisie.

Pierwsza z kategorii opisuje metody, których kluczowym elementem jest szukanie różnicy pomiędzy globalnymi cechami dwóch chmur punktów 3D. Do tej kategorii należą opisane wcześniej PointNetLK, PCRNNet oraz dodatkowo Feature-metric Registration. Feature-metric Registration [100] jest metodą koncentrującą się na minimalizowaniu różnic cech między dwiema chmurami punktów zamiast na wyszukiwaniu korespondencji pomiędzy punktami. Metoda korzysta z dwóch modułów: enkodera do opisu charakterystycznych cech chmury punktów oraz wielozadaniowy moduł do nienadzorowanego uczenia oraz estymowania finalnej transformacji. Architektura korzysta z uczenia częściowo nadzorowanego, co pozwala sieci na naukę ekstrakcji charakterystycznych, globalnych cech wykorzystywanych w dopasowywaniu chmur punktów.

Druga, zdecydowanie liczniejsza kategoria polega na ustalaniu korespondencji i estymowaniu transformacji. Metody te udoskonalają klasyczne podejście do dopasowywania chmur punktów, które można w skrócie opisać za pomocą trzech kroków: znajdowanie punktów charakterystycznych oraz opis cech, porównywanie i odrzucanie błędnych dopasowań, estymacja transformacji. Do tej kategorii należą również rozwiązania skupiające się na jednym z wymienionych kroków. Poza wcześniej opisanymi metodami takimi jak: 3DMatch, 3DFeat-Net, PPFNet, DCP do tej kategorii zostały sklasyfikowane także:

- Ekstraktory cech: 3D Local Feature, FCGF,
- Metody znajdujące dopasowania: DeepVCP,
- Metody odrzucające wartości odstające: 3DRegNet, DSAC, Unsupervised Consensus Maximization, Deep Global Registration.

3D Local Feature [101] przedstawia kompletny proces dopasowywania chmur punktów 3D. Ideą tej metody jest to, że lokalny deskryptor, para punktów, która daje dobre rezultaty w dopasowaniu chmur punktów jest też dobrym punktem startowym do estymacji lokalnej orientacji. Metoda rozszerza architekturę PPF-FoldNet poprzez dodanie głęboko wyuczonej

orientacji. Dzięki temu możliwe jest poznanie orientacji obiektu niezależnie od ich rozmieszczenia przestrzennego. Bazując na tym autorzy zaproponowali schemat hypothesize-and-verify, który na podstawie lokalnych, dobrych dopasowań estymuje finalną transformację. Aby tego dokonać, autorzy stworzyli wyspecjalizowaną architekturę sieci (RelativeNet), która dla każdego z punktów charakterystycznych estymuje relatywną pozę. FCGF (Fully Convolutional Geometric Features [102]) jest architekturą w pełni wykorzystującą konwolucyjne geometryczne cechy, dzięki zastosowaniu tensorów (sparse tensor representation) umożliwiającym efektywne przechowywanie w pamięci klasycznych danych wejściowych konwolucyjnej sieci neuronowej. Autorzy zaproponowali również konwolucyjną metrykę do uczenia, ponieważ w tym przypadku występuje lokalna korelacja pomiędzy sąsiednimi cechami. Dodatkowo liczba cech w pełni konwolucyjnej architekturze jest zdecydowanie większa niż w klasycznym podejściu np. liczenia dystansów dla każdego fragmentu danych. Architektura FCGF jest jednocześnie bazą dla dwóch innych metod: DGR (Deep Global Registration [103]) oraz Learning multiview 3D point cloud registration (LMVD [104]). Pierwsza z nich opiera się na trzech modułach: 6-wymiarowej sieci konwolucyjnej do znajdowania korespondencji, różniczkowalnym algorytmie Procrusta oraz bazujący na gradientach $SE(3)$ [105] optymalizator do estymacji transformacji. Learning multiview 3D point cloud registration jest metodą łączącą naukę zgrubnego oraz dokładnego dopasowania w jeden proces. Polega na rozwiązywaniu różniczkowalnego, optymalizacyjnego problemu Procrusta związanego z estymacją parametrów transformacji pomiędzy parami chmur punktów. Wartością dodaną jest warstwa sieci (overlap pooling) przewidująca poziom ufności (confidence) względem parametrów transformacji. DeepVCP jest metodą korzystającą z architektury PointNet++. Bazuje na uczeniu wykrywania punktów charakterystycznych oraz oryginalnej metodzie wykrywania korespondencji oraz funkcji strat biorącej pod uwagę lokalne podobieństwo oraz estymację globalnej transformacji. Alternatywami dla klasycznego algorytmu RANSAC są DSAC (Differentiable RANSAC) [106] oraz jego nowsza wersja przedstawiona w artykule "Learning Less is More – 6D Camera Localization via 3D Surface Regression"[107]. Autorzy przedstawiają oryginalne podejście do wyszukiwania pozycji kamery. Za pomocą konwolucyjnej sieci neuronowej estymowane są korespondencje 2D-3D, z których wybierany jest podzbiór czterech współrzędnych sceny. Dla każdej tak zdefiniowanej hipotezy tworzony jest obraz reprojekcji błędów, który jest oceniany przez kolejną konwolucyjną sieć neuronową. Na podstawie rozkładu otrzymanych wyników

wybijana jest ostateczna pozycja kamery. Autorzy przedstawiaj DSAC jako komponent, ktry w przyszoci moe zosta wdroony w dowoln architektur opart na gebokim uczeniu. Unsupervised Consensus Maximization jest metod, ktra moe zosta z powodzeniem uyta w dopasowywaniu chmur punktw. Poprzez wykorzystanie sieci (PointNet segmentation architecture) oraz macierzy Vandermonde’a [108], autorzy pokazuj jak funkcja straty Vandermonde’a moe zosta zaadoptowana do problemw zwizanych z transformacj 3D. Inn architektur, ktra w skuteczny sposb odrzuca wartoci odstajce, tj. bedne korespondencje pomidzy dwoma skanami, jest 3DRegNet [109]. Autorzy proponuj architektur skadajc si z moduu klasyfikacji oraz dopasowywania. Pierwszy z nich ze zbioru wszystkich korespondencji estymuje wyniki, wagi, na podstawie ktrych wskazywane s poprawne oraz bedne dopasowania. Modu dopasowania generuje finaln transformacj.

Kolejn prb podziau metod gebokiego uczenia podjto w nowym, jeszcze nieoficjalnie opublikowanym artykule ”Deep Learning vs. Traditional 3d Registration: A Featureless 3d Registration Baseline”[110]. Metody zosta sklasyfikowane na trzy kategorie:

- Feature learning, czyli uczenie charakterystycznych cech, midzy innymi detekcja punktw charakterystycznych oraz ich opis za pomoc deskryptorw,
- Robust estimation learning, zbierajca metody skupiajce si na odrzucaniu bednych, odstajcych wartoci ze zbioru cech lub korespondencji,
- End-to-end, opisujca kompleksowe metody uczce caego procesu aczenia chmur punktw 3D.

Pierwsza z nich obejmuje dwie podkategorie, metody uczce si detekcji punktw charakterystycznych oraz metody wykorzystywane do opisu cech. Poza wczeniej opisanymi metodami (3DMath, 3DFeat-Net) do pierwszej podkategorii nale: D3Feat, USIP, Predator. D3Feat [111] wykorzystuje w peni konwolucyjn sie opart na KPConv [112] do jednoczesnego uczenia detekcji i opisu lokalnych cech 3D. Istotnym elementem metody jest nowatorska strategia wyboru punktw charakterystycznych, ktra zapewnia powtarzalno wynikw nawet w przypadku chmur punktw 3D o zmiennej geoci. Dodatkowo, autorzy proponuj wykorzystanie funkcji strat bazujcej na informacjach pochodzcych z procesu uczenia porwnywania cech w module samonadzorowanego detektora. USIP (Unsupervised Stable Interest Point [113]) jest metod bazujc na autorskiej architekturze FPN (Feature Proposal Network), ktra generuje potencjalne punkty kluczowe w przeciwiestwie do

wybierania już istniejących na chmurze punktów. Predator (Pairwise point cloud registration with deep attention to the overlap region [114, 115]) jest architekturą sieci neuronowej uczącą się wykrywać obszar wspólny pomiędzy dwiema chmurami punktów. Posiada zdolność do priorytetyzowania punktów/fragmentów istotnych dla procesu dopasowywania, co pozytywnie wpływa na jego efektywność. Autorzy przedstawili w swojej pracy także schemat udoskonalania deskryptorów punktów charakterystycznych, poprzez aktualizowanie informacji pochodzących z dopasowywanej chmury. Dodatkowo, zaproponowano nowatorską funkcję strat do uczenia, która pomaga w próbkowaniu i wykrywaniu bardziej powtarzalnych punktów charakterystycznych.

Do podkategorii deskryptorów zaklasyfikowano takie metody jak opisane wcześniej: FCGF, PPFNet oraz SpinNet, GeDi, DIP, Imfnet, CGF, YOHO, LMVD. SpinNet [116] to sieć neuronowa składające się z dwóch modułów: Przestrzennego Transformatora Punktowego (Spatial Point Transformer), który przekształca wejściowe skany 3D w starannie zaprojektowaną przestrzeń cylindryczną, aby uzyskać niezależność od rotacji oraz Ekstraktora Cech Neuronowych (Neural Feature Extractor), który wykorzystuje konwolucyjne warstwy neuronowe, aby nauczyć się unikalnych wzorców lokalnych cech. GeDi [117] jest metodą opierającą się na opisie nie pojedynczego punktu 3D, ale fragmentu danych jako zbioru sąsiadujących ze sobą punktów. Dla tak wybranego fragmentu obliczany jest lokalny układ odniesienia (LRF), który ma na celu wyznaczenie jego kanonicznej reprezentacji [118]. Do tego procesu wykorzystywany jest QNet (sieć transformacji kwaternionowej inspirowanej artykułem [119]). Taka reprezentacja fragmentu danych jest procesowana przez punktową głęboką sieć neuronową niewrażliwą na permutację w celu wygenerowania deskryptora. CGF (Compact Geometric Features [120]) jest deskryptorem cech wykorzystującym nieuporządkowane chmury punktów, nie wymagając przy tym spójnej topologii powierzchni, wolumetrycznej reprezentacji chmury punktów ani syntezy multimodalnej z innym typem danych np. z obrazami głębi. Dzięki temu, że wyliczany deskryptor posiada wsparcie do wyszukiwania najbliższych sąsiadów w przestrzeni Euklidesowej, korespondencje wyznaczone są bardzo szybko i efektywnie. Autorzy opisują metodę jako wymienny komponent, który może zostać wykorzystany w dowolnych procesach wykorzystujących klasyczne deskryptory. DIP (Distinctive 3D Local Deep Descriptors [121]) to metoda polegająca na ekstrakcji fragmentów chmur punktów, umieszczania ich w lokalnym układzie odniesienia i kodowaniu ich za pomocą głębokiej sieci neuronowej opartej na PointNet.

Architektura jest uczona całego, kompleksowego procesu dopasowywania na lokalnych i losowo wybranych punktach, co znacząco poprawia skuteczność i wyniki nawet dla danych nie uwzględnionych podczas uczenia. Deskrytory te kodują tylko lokalną informację geometryczną na podstawie zbioru punktów, co sprawia, że są odporne na zaszumienie, zakrycia i brakujące obszary. YOHO (You Only Hypothesize Once [122]) jest nowatorską metodologią do łączenia chmur punktów wykorzystującą oraz analizującą orientacje lokalnych podzbiorów punktów w celu wygenerowania deskryptora oraz znalezienia globalnego dopasowania. Autorzy szczerzą się zautomatyzowanym procesem radzącym sobie nawet w przypadku dużego szumu pomiarowego oraz przy chmurach punktów posiadających bardzo różne gęstości punktów. Kolejną, kompleksową metodę dopasowywania chmur punktów na podstawie deskryptora zaproponowano w artykule "End-to-End Learning Local Multi-view Descriptors for 3D Point Clouds [123]". Wykorzystana architektura sieci składa się z trzech etapów: renderowanie wielu wirtualnych widoków dla punktu charakterystycznego, wyznaczenie cech każdego z widoków, fuzja informacji pochodzących ze wszystkich widoków. Autorzy posłużyli się wbudowanym w sieć różniczkowalnym mechanizmem renderowania [124] w celu projektowania lokalnej geometrii 3D punktów na obraz 2D. Finalnym opisem cechy zajmuje się moduł sieci łączący informację ze wszystkich widoków za pomocą próbkowania wektorów cech. IMFNet (Interpretable Multimodal Fusion Network [125]) jest multimodalną metodą łączącą informację o strukturze chmury punktów, jak i informację o teksturze odpowiadającego jej obrazu. IMFNet składa się z dwóch głównych modułów: modułu fuzji multimodalnej i nowatorskiego modułu DAM (Descriptor Activation Mapping). Moduł fuzji multimodalnej oparty jest na architekturze kodera i dekodera, które są wzorowane na opisanym wcześniej FCGF [102]. Wykorzystując moduł uwagi inspirowany transformerem [126], IMFNet ekstrahuje ważne informacje o teksturze dla każdego punktu. Następnie dane dotyczące struktury i tekstury są łączone i przekazywane do modułu dekodera w celu nauki ostatecznego deskryptora. Drugi z modułów (DAM) umożliwia interpretację wpływu sąsiadujących punktów w proces wyznaczania deskryptora.

Druga kategoria (Robust estimation learning) obejmuje metody, które zamiast wyznaczać lepsze opisy cech do procesu dopasowania skupiają się na efektywnym odrzucaniu błędnych dopasowań ze zbioru cech lub oszacowanych korespondencji. Wśród tych metod są opisane wcześniej 3DRegNet, DGR oraz takie metody jak: DHVR, PointDSC, CSCE-Net, TEASER, SC2-PCR, VRNet oraz TR-DE. Metodologia DHVR (Deep Hough Voting for Robust Global

Registration [127]) wykorzystuje trójki (ang. triplets) korespondencji w celu znalezienia najlepszej transformacji z sześciowymiarowej przestrzeni Hough'a [128]. PointDSC (Deep Spatial Consistency [129]) jest dwu-etapową metodą bazującą na zgodności przestrzennej (ang. 3D spatial consistency). Idea ta jest wykorzystywana również w tradycyjnym, dobrze znanym algorytmie SM (Spectral Matching [130]). Polega ona na zachowaniu odległości pomiędzy parami punktów 3D przy przekształceniu estymowaną transformacją. Pierwszy etap obejmuje wyznaczenie cech, drugi ich filtrację poprzez odrzucenie błędnych wartości. Inną metodą również wykorzystującą zgodność przestrzenną jest architektura CSCE-Net (Channel-Spatial Contextual Enhancement Network [131]). Poprzez wykorzystanie mechanizmu opartego o informacje przestrzenne, metoda skupia się na ważnych regionach chmury punktów, podczas gdy drugi mechanizm akcentuje informację pochodzącą z innych kanałów lub analizy cech. Ich połączenie daje pełne zrozumienie danych oraz potencjalne korespondencje, z których błędne dopasowania są odrzucane za pomocą informacji o zgodności przestrzennej. TEASER (Truncated least squares Estimation And SEMidefinite Relaxation [132, 133]) formułuje problem dopasowania za pomocą funkcji kosztu opartej o TLS (Truncated Least Squares [134]), który jest rozwiązywany za pomocą wejściowych korespondencji. Dodatkowo, estymowanie transformacji następuje za pomocą jej dekompozycji na skalę, rotację i translację. VRNet ([135]) jest architekturą sieci opierającą się na uczeniu macierzy dopasowań pomiędzy zaproponowanymi przez autorów cechami punktów RCP (Rectified Corresponding Points). Finalna macierz transformacji jest estymowana za pomocą rozwiązywania problemu Procrusta. TR-DE (Point Cloud Registration via Novel Transformation Decomposition [136]) ze zbioru korespondencji odrzuca błędne dopasowania poprzez rozkład problemu transformacji z sześcioma stopniami swobody (ang. 6 DOF, degrees of freedom) na dwa pomniejsze rozwiązywalne zagadnienia (2+1) DOF i (1+2) DOF. Pierwszym z nich jest znalezienie (2+1) stopni swobody, gdzie liczba dwa reprezentuje rotację z dwoma stopniami swobody, a jedynka translację wzdłuż wektora rotacji. Analogicznie, w (1+2) DOF, jedynka reprezentuje kąt rotacji, a dwójka przemieszczenie prostopadłe do wektora rotacji z (2+1) DOF. SC2-PCR (Second order spatial Compatibility for Point Cloud Registration [137]) jest metodologią badającą miarę zgodności przestrzennej drugiego rzędu, aby obliczyć podobieństwo między korespondencjami. Warto zaznaczyć, że nie jest to metoda oparta na głębokim uczeniu, ale często pojawia się w porównaniach i wynikach innych publikowanych artykułów ze względu na swoją skuteczność.

Ostatnia kategoria obejmuje kompleksowe rozwiązania, które polegają nie tylko na uczeniu opisu cech charakterystycznych, ale również następnego kroku wymaganego do poprawnego połączenia chmur punktów jakim jest dopasowywanie cech oraz znajdowanie korespondencji. Również i w tym wypadku, metody można sklasyfikować na mniejsze podgrupy. Pierwsza z nich obejmuje metody iteracyjne, działające na podstawie idei ICP. Do tej kategorii należą opisane wcześniej DCP oraz sieć PRNet. DCP jest pionierem dla późniejszych rozwiązań w tej kategorii. Dodatkowo, sklasyfikowano tutaj takie metody jak: RPM-Net, RGM, Stickypillars, IDAM, PCAM, DeepGMR. RPM-Net (Robust Point Matching Network [138]) to metoda oparta na różniczkowalnej głębokiej sieci neuronowej. Wykorzystuje korespondencje punktów oparte na hybrydowych cechach zamiast na odległościach przestrzennych. Dodatkowo, korzysta z warstwy Sinkhorna [139] i procesu wyżarzania (ang. annealing [140]) do uzyskania wstępnego przyporządkowania korespondencji. RPM-Net iteracyjnie doskonali łączenie chmur punktów oraz posiada dodatkową sieć przewidującą optymalne parametry annealingu. RGM (Robust Point Cloud Registration Framework Based on Deep Graph Matching [141]) poza porównywaniem cech, korzysta ze strukturalnej informacji pochodzącej z grafów w celu wyeliminowania większej ilości błędnych dopasowań. Stickypillars [142] wykorzystuje sieć GNN (Graph Neural Network [143]) oraz architekturę transformera wraz z mechanizmem uwagi [126], aby zebrać informację o punktach charakterystycznych 3D. Pozwala to dokładnie dopasować cechy dwóch chmur punktów, bez wykorzystania deskryptorów lub innych heurystycznych strategii dopasowywania. IDAM (Iterative Distance-Aware Similarity Matrix Convolution with Mutual-Supervised Point Elimination for Efficient Point Cloud Registration [144]) jest modelem wykorzystującą iteracyjny moduł splotu macierzy podobieństwa z uwzględnieniem opisu cech geometrycznych oraz odległości w przestrzeni euklidesowej. Ponadto, autorzy zaproponowali dwuetapową metodę eliminacji punktów w celu usunięcia błędnych dopasowań oraz nowatorską funkcję straty do nadzoru uczenia. PCAM (Product of Cross-Attention Matrices for Rigid Registration of Point Clouds [145]) to architektura wykorzystująca macierze wzajemnej uwagi (ang. cross-attention matrices) do łączenia lokalnej, niskopoziomowej informacji geometrycznej z informacją z jakiego fragmentu chmury punktów pochodzi dany punkt charakterystyczny. Poprzez tak sformułowane dopasowania oraz wymianę informacji na wielu poziomach sieci, metoda skutecznie dopasowuje punkty, jednocześnie eliminując fałszywe dopasowania. DeepGMR (Deep Gaussian Mixture

Registration [146]), jest metodą łączącą Gaussian Mixture Model (GMM [147]) z sieciami neuronowymi. Metoda formułuje problem dopasowania chmur punktów jako minimalizacja rozbieżności (Dywergencja Kullbacka-Leiblera [148]) pomiędzy rozkładami prawdopodobieństwa reprezentowanymi przez modele rozkładu mieszanin Gaussowskich.

Kolejna podgrupa reprezentuje metody wykorzystujące opis globalnych wektorów cech do łączenia chmur punktów. Opierają się one w głównej mierze na architekturze PointNet. Są to opisane wcześniej PointNetLK, PCRNet, Feature-metric Registration oraz dodatkowo OMNet. OMNet (Learning Overlapping Mask for Partial-to-Partial Point Cloud Registration [149]) jest iteracyjną metodą bazującą na wzajemnie nakładających się fragmentach chmur punktów. W każdym kroku metoda uczy się masek opisujących wspólne fragmenty. Ich głównym celem jest identyfikacja oraz odrzucenie nienakładających się punktów podczas łączenia chmur punktów. Metoda stopniowo udoskonala oszacowanie transformacji poprzez iteracyjne dostosowywanie nakładających się masek i aktualizowanie cech globalnych.

Ostatnia, najmniej liczna podkategoria obejmuje metody korzystające z mechanizmów samouwagi (ang. self-attention) i wzajemnej uwagi (ang. cross-attention) w celu znalezienia finalnej transformacji za pomocą dopasowanych tzw. super punktów. Zamiast niezależnego rozpatrywania poszczególnych punktów, superpunkty są tworzone przez agregację sąsiednich punktów, które mają wspólne cechy lub wykazują spójność przestrzenną. Grupa metod reprezentująca tę podkategorię to: REGTR, GeoTransformer, GLORN, Leopard. REGTR (End-to-end Point Cloud Correspondences with Transformers [150]) jest metodą, która zamiast dopasowywania cech charakterystycznych i przeprowadzania iteracyjnego procesu estymacji transformacji, stosuje mechanizmy uwagi do bezpośredniego przewidywania zbioru korespondencji między chmurami punktów. Architektura sieci składa się głównie z warstw transformera [126], które zawierają mechanizmy samouwagi i wzajemnej uwagi, umożliwiając sieci naukę bezpośredniego przewidywania korespondencji. GeoTransformer (Geometric Transformer for Fast and Robust Point Cloud Registration [151]) wykorzystuje wyuczone cechy geometryczne w celu dokładnego dopasowywania pomiędzy superpunktami, które są opisywane za pomocą odległości pomiędzy parami punktów oraz kątami pomiędzy trójkami (tripletami) punktów. Wykorzystuje grupę hiperparametrów do kontrolowania czułości zmian odległości i kątów, które muszą być precyzyjnie wybrane dla różnych zestawów danych. Dodatkowo uwzględniane są obszary wspólne dwóch chmur punktów, w celu ponownego ważenia każdego dopasowania superpunktów na podstawie stosunku nakładających się

fragmentów. GLORN [152] korzysta z pełnej sieci konwolucyjnej do wyszukiwania superpunktów, koncentrując się na obszarach wspólnych chmur punktów. Dodatkowo w wybranych superpunktach generowany jest deskryptor oraz odrzucane są punkty nie należące do nakładających się fragmentów. Deskryptory cech generowane przez tę sieć są niezależne od rotacji. Leopard [153] w celu uchwycenia informacji zakodowanych w cechach geometrycznych oraz informacji o relatywnych pozycjach punktów dzieli reprezentację chmury punktów na dwie przestrzenie: przestrzeń cech oraz przestrzeń pozycji 3D. Metoda kodowania pozycji poprawia zrozumienie przez model relacji przestrzennych, podczas gdy moduł repozycjonowania dodatkowo poprawia dopasowanie między chmurami punktów.

Ciekawym źródłem dostępnych licznych metod dopasowywania chmur punktów okazało się publiczne repozytorium [154]. Jest to zestawienie metod, narzędzi i zbiorów danych związanych z procesem łączenia chmur punktów. Lista znajdujących się tam metod zostało pogrupowana na metody łączenia wstępnego oraz dokładnego (np. opisane wcześniej ICP, DeepICP). Metody łączenia zgrubnego podzielono na następujące kategorie:

- Feature Matching Based - ogólna kategoria klasyfikująca metody mające udział w procesie łączenia chmur punktów: detekcja punktów kluczowych, opis cech, odrzucanie błędnych dopasowań oraz dodatkowo metody oparte na grafach.
- End-to-End - kategoria opisująca kompleksowe metody estymujące finalną transformację pomiędzy chmurami punktów,
- Randomized - zawiera metody takie jak RANSAC czy 4PCS [155]/Super 4PCS [156].
- Probabilistic - zawiera metody między innymi oparte na metodach mieszanin Gaussa, jak na przykład DeepGMR oraz metody probabilistyczne.
- Others - pozostałe metody nie należące do innych kategorii.

W wymienionym repozytorium znajdują się ogromna liczba metod, dokładnie 187. W dalszej części przedstawiono po kilka najnowszych metod, z każdej z podgrup opartych na głębokim uczeniu, które nie zostały dotąd opisane. Pominięto grupę metod Randomized, gdyż nie są one metodami opartymi na sieciach neuronowych.

Pierwszymi trzema metodami opisującymi ekstrakcję cech, czyli szukanie punktów charakterystycznych lub ich opis, są: MaskNet, UPPNet, ImLoveNet. MaskNet [157] jest architekturą uczącą się znajdować poprawne dopasowywania poprzez estymację oraz

nałożenie maski, która zachowuje jedynie te punkty, które jak najlepiej aproksymują swoich przedstawicieli z dopasowywanej chmury. Jest to metoda znajdująca zastosowanie we wstępnym przetwarzaniu - zwiększa skuteczność innych metod łączenia chmur punktów poprzez znalezienie części wspólnej źródłowej oraz dopasowywanej chmury punktów. UPPNet [158] jest architekturą, która najpierw przewiduje, które regiony pomiędzy dwiema chmurami punktów się nakładają, a następnie wykorzystuje moduł znajdowania superpunktów oraz moduł tzw. udoskonalania (ang. fine-grained refinement) w celu precyzyjnej identyfikacji poprawnych korespondencji. ImLoveNet [159], podobnie jak PREDATOR, jest metodą wykorzystywaną do łączenia chmur punktów posiadających bardzo małą część wspólną za pośrednictwem zwykłego obrazu RGB przedstawiającego rzeczywisty obraz wspólnego fragmentu obu chmur.

Ciekawą alternatywę, znajdującą poprawne dopasowania pomiędzy punktami proponuje DetarNet. DetarNet [160] dzieli transformację na dwie składowe, rotację oraz translację. Składowa translacji jest znajdowana przez zaproponowany moduł oparty o architekturę sieci syjamskiej PCFD (Progressive and Coherent Feature Drift). Kolejnym krokiem jest wykorzystanie modułu CEU (Consensus Encoding Unit), w celu wyznaczenia rozpoznawalnych cech dla podzbioru domniemanych korespondencji. Ostatnim krokiem jest budowa sieci klasyfikacji (SCA, Spatial and Channel Attention) w celu znalezienia poprawnych odpowiedników oraz wykorzystanie metody opartej o rozkład wartości osobliwych (SVD) do wyznaczenia rotacji. Quatro [161] jest metodą, która mierzy się z problemem zbyt małej ilości poprawnych dopasowań pomiędzy chmurami reprezentującymi środowiska miejskie. Ponieważ często źródłowa chmura oraz docelowa są rejestrowane w innej lokalizacji, to fragmenty reprezentujące lokalnie tę samą przestrzeń mogą mieć różne zagęszczenie.

UTOPIC (UncerTainty-aware Overlap PredICtion [162]) jest kompleksową metodą rozwiązującą problem niejednoznacznego przewidywania obszarów nakładających się pomiędzy chmurami punktów. Metoda korzysta z transformera w celu uzyskania cech niezależnych od transformacji oraz precyzyjnie opisujących lokalną geometrię.

LSG-CPD [163] reprezentuje grupę metod probabilistycznych. Łączy ona metodę CPD (Coherent Point Drift [164]) z informacją o lokalnej geometrii płaszczyzny, poprzez jej zmienność (krzywizna) oraz wektory normalne. W zależności od geometrii, metoda generuje odpowiednie komponenty GMM (Gaussian Mixture Model). Następnie formułowany jest

problem dopasowania chmur punktów jako problem oszacowania największej wiarygodności (Maximum Likelihood Estimation, MLE [165]), który jest rozwiązywany za pomocą algorytmu maksymalizacji oczekiwań (Expectation-Maximization, EM [166]).

DeepUME [167] jest architekturą opierającą się o metodę UME (Universal Manifold Embedding [168]) do dopasowywania chmur punktów 3D. Składa się ona z trzech modułów. Pierwszy z nich konstruuje trójwymiarowy układ współrzędnych przy użyciu PCA (Principal component analysis [169]), drugi implementuje sieć neuronową do próbkowania surowych chmur punktów 3D, trzeci blok, implementuje metodę UME.

Jeszcze innym źródłem interesujących metod łączenia chmur punktów okazała się strona internetowa [170] grupująca oraz porównująca metody dopasowujące chmury punktów. Strona posiada ranking metod najlepiej realizujących dany przypadek łączenia chmur punktów na określonym zbiorze danych. Ranking śledzi progres rozwoju metod dopasowywania chmur punktów. Do metod opisanych dotychczas warto dodać: MS-SVConv, 3DSmoothNet, R-PCR, GCNet. Multi-Scale Sparse Voxel Convolution (MS-SVConv [171]) jest metodą opartą na architekturze sieci U-Net [172] do liczenia deksyptorów chmur punktów 3D. MS-SVConv korzysta z silników do liczenia konwolucji, takich jak MinkowskiNet [102] lub SPVNAS [173]. W przeciwieństwie do innych metod opartych na U-Net, MS-SVConv generuje deksyptory, które mają zdolność do efektywnego działania na zupełnie nowych zestawach danych poprzez wykorzystanie metody TL (Transfer Learning [174]). 3DSmoothNet [175] to metoda opierająca się na architekturze syjamskiej uczenia głębokiego. Wykorzystuje wokselową reprezentację wartości SDV (Smoothed Density Value), która jest obliczana dla każdego punktu charakterystycznego. SDV jest następnie dopasowywane do lokalnego układu odniesienia (LRF), co pozwala na osiągnięcie niezależności względem obrotu. Deskryptor wygenerowany przez 3DSmoothNet jest kompaktowy, składa się zaledwie z 32 wymiarów wyjściowych. Główną zaletą tej metody jest nisko wymiarowy deskryptor cech (32 wymiary wyjściowe), co pozwala na zastosowania w czasie rzeczywistym. Jedną z najnowszych metod jest R-PCR (Recurrent Point Cloud Registration [176]). To nowatorska architektura głębokiej sieci neuronowej, która rozwiązuje problem łączenia chmur punktów poprzez skuteczne łączenie globalnych cech oraz integrację wysokiego rzędu procesu decyzyjnego Markowa [177] w iteracyjnym procesie dopasowywania punktów. Architektura R-PCR składa się z kilku kluczowych elementów. Wykorzystuje architekturę PointNet do ekstrakcji globalnych informacji o geometrii z chmur punktów źródłowych i docelowych. Do efektywniejszego

procesu dopasowywania, R-PCR wprowadza moduł krzyżowy oraz sieć neuronową do łączenia informacji pomiędzy parami chmur punktów. Dodatkowo, wykorzystuje rekurencyjny operator aktualizacji oparty na GRU (Gate Recurrent Unit [178]), aby iteracyjnie aktualizować oszacowaną transformację i uwzględniać informacje z poprzednich kroków. GCNet (Geometry-guided Consistent Network [179]) jest metodą pomagającą odrzucać błędne dopasowania poprzez wykorzystanie wewnętrznych proporcji korespondencji. Jest to wykonywane za pomocą dekodera hierarchii w kształcie piramidy, spójnej strategii głosowania do otrzymywania poprawnych korespondencji oraz modułu kodowania uwzględniającego cechy geometryczne.

2.2 Podsumowanie

Wybór odpowiedniej metody łączenia chmur punktów ma kluczowe znaczenie dla osiągnięcia optymalnej wydajności procesu. Wybór ten w dużej mierze zależy od czynników takich jak charakterystyka chmury punktów, problem z jakim mamy do czynienia oraz pożądana dokładność dopasowania. W tym rozdziale przedstawiono bardzo różne rozwiązania posiadające swoje mocne i słabe strony. W tabeli 2 zestawiono metody wykorzystywane w klasycznym procesie łączenia chmur punktów.

Niestety, metody łączenia chmur punktów oparte na głębokim uczeniu są na stosunkowo wczesnym etapie rozwoju, dlatego uzyskanie kompletnych wyników i porównania ze wszystkich opisanych dotąd algorytmów jest trudne. Wiele z nich odnosi się do specyficznych danych lub porównywane są jedynie z kilkoma innymi algorytmami. Dopiero nowsze publikacje starają się uwzględniać w swoich testach dane z innych źródeł niż własne. Takim głównym benchmarkiem, zbiorem danych na którym są testowane metody głębokiego uczenia jest 3DMatch benchmark [180]. Analiza artykułów dotyczących metod głębokiego uczenia wraz artykułami przeglądowymi w kontekście łączenia chmur punktów 3D pozwala je zgrubnie sklasyfikować na trzy kategorie: opis cech, odrzucanie wartości błędnych, metody kompleksowe. Zestawienie metod głębokiego uczenia zostało zgrupowane w trzy tabele. Pierwsza z nich opisuje deskryptory na podstawie miary Feature Matching Recall (FMR) - jest to miara statystyczna, która oblicza stosunek pozytywnych dopasowań znalezionych przez algorytm do całkowitej liczby możliwych poprawnych dopasowań. Kolejną miarą jest Registration Recall (RR) - jest to stosunek poprawnie dopasowanych par chmur punktów do całego zbioru danych. Poprawnie dopasowane chmury (dla zbioru 3DMatch) są to pary

PUNKTY CHARAKTERYSTYCZNE	DESKRYPTORY	ODRZUCANIE BŁĘDNYCH DOPASOWAŃ	DOKŁADNE DOPASOWYWANIE
<ul style="list-style-type: none"> - SIFT3D [45, 46] - SURF3D [47, 48] - ISS [49] - HARRIS3D [50] 	<ul style="list-style-type: none"> - SI [51] - PFH [52] - FPFH [52] - VFH [53] - SHOT [54] - 3DSC [51] - USC [54] - CORS [55] - MI [56] - NBS [57] - PCE [58] - HWR [47, 60] - SURF/USURF [47, 60] - PIG [61] - HOG [61] - RIFT [62] - PFHRGB [63] - SHOTColor [54] - ISE [64] 	<ul style="list-style-type: none"> - RANSAC [44] 	<ul style="list-style-type: none"> - ICP [39, 40]

Tabela 2: Zestawienie algorytmów klasycznego procesu łączenia chmur punktów.

posiadające co najmniej 30% wzajemnego pokrycia oraz błąd dopasowań (RMSE, Root Mean Square Estimation) losowo próbkowanych punktów niższy niż 0.2[m]. Miarą RR można opisywać metody kompleksowe oraz metody odrzucające błędne dopasowania.

W tabeli 3 przedstawiono wyniki FMR dla grupy metod opisujących cechy chmur punktów opartych na głębokim uczeniu oraz cztery klasyczne deskryptory.

Metoda	Feature Matching Recall (%)
SI [51]	22.7
SHOT [54]	23.8
FPFH [52]	35.9
USC [54]	40
PointNet [74]	47.1
CGF [120]	58.2
PPFNet [80]	62.3
3DMatch [76]	66.8
PPF-FoldNet [81]	71.8
3D Local Feature [101]	74.6
3DSmoothNet [175]	94.7
FCGF [102]	95.2
D3Feat [111]	95.8
Predator [114]	96.6
DIP [121]	96.8
LMVD [104]	97.5
SpinNet [116]	97.6
GeDi [117]	97.9
YOHO [122]	98.2
MS-SVConv [171]	98.4
IMFNet [125]	98.6

Tabela 3: Zestawienie wyników FMR dla różnych metod opisujących cechy.

Kolejna tabela podsumowuje metody, które skupiają się na odrzuceniu błędnych dopasowań. Niestety i w tym przypadku część metod jest testowana na innym zbiorze danych, korzysta z innego modelu lub innego deskryptora cech. W tabeli 4 znajduje się zestawienie metod z ich najlepszymi wynikami wykorzystującymi deskryptory: FPFH, FCGF albo 3DSmoothNet.

Metody	Registration Recall (%)
PointNetLK [86, 87]	1.61
DCP [92]	3.22
VRNet [135]	72.9
3DRegNet [109]	77.76
DetarNet [160]	84.9
DGR [103]	88.85
DHVR [127]	91.93
TEASER [132, 133]	91.99
PointDSC [129]	92.85
TR-DE [136]	92.98
SC2-PCR [137]	93.28
CSCE-Net [131]	93.47

Tabela 4: Zestawienie wyników RR dla metod odrzucających błędne dopasowania.

Tabela 5 jest zestawieniem metod opisujących cechy z wykorzystaniem algorytmu RANSAC.

Metoda	Registration Recall (%)
3DSmoothNet [175]	81.5
D3Feat [111]	81.6
FCGF [102]	85.1
SpinNet [116]	88.6
Predator [114]	89
CoFiNet [181]	89.3
YOHO [122]	90.5
GCNet [179]	92.9
Lepard [153]	93.5

Tabela 5: Zestawienie wyników RR dla metod opisujących cechy (RANSAC).

W ostatniej tabeli 6 zestawiono metody kompleksowe, które osiągnęły bardzo dobre wyniki.

Metoda	Registration Recall (%)
PointNetLK [86, 87]	1.61
DCP [92]	3.22
OMNet [149]	35.9
DGR [103]	85.3
PCAM [145]	85.5
RegTR [150]	93
GeoTransformer [151]	95

Tabela 6: Zestawienie wyników RR dla kompleksowych metod.

Pozostałe metody nie zostały uwzględnione w zestawieniach z jednego, bądź kilku wymienionych poniżej powodów:

- Testy uwzględniały zupełnie inne zestawy danych lub modele.
- Testy skupiały się głównie na wybranym algorytmie analizując skuteczność w kontekście doboru jego parametrów, a nie skuteczności łączenia.
- Opisywana metoda nie służyła bezpośrednio do łączenia chmur punktów.

3 Charakterystyka danych wejściowych

Dane 3D wykorzystane w niniejszej rozprawie są wynikiem dwóch projektów rewitalizacyjnych oraz dokumentacyjnych przeprowadzanych w Muzeum Pałacu Króla Jana III w Wilanowie:

- Antygabinet Królowej i Gabinet al Fresco - dokumentacja w ramach projektu „Restauracja i zabezpieczenie symbolu polskiego dziedzictwa kulturowego - Muzeum Pałacu Króla Jana III w Wilanowie” współfinansowany przez Unię Europejską ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Infrastruktura i Środowisko,
- Gabinet Chiński Króla (pomiar po zakończeniu prac konserwatorskich) i Garderoba Króla – dokumentacja w ramach projektu „Rewitalizacja i digitalizacja jedynej w Polsce barokowej rezydencji królewskiej w Wilanowie” (Program Operacyjny Infrastruktura i Środowisko) finansowany ze środków Unii Europejskiej i Ministerstwa Kultury i Dziedzictwa Narodowego.

3.1 Opis skanowanych pomieszczeń zabytkowych

Pałac w Wilanowie jest jedyną barokową rezydencją królewską w Polsce. Budowę tej letniej rezydencji rozpoczęto pod koniec XVII wieku, a następnie wielokrotnie ją rozbudowywano i modernizowano. Dwukondygnacyjny budynek pałacu kryje w swoim wnętrzu wiele pomieszczeń o stosunkowo niewielkich rozmiarach, ale za to charakteryzujących się bardzo bogatym i zróżnicowanym wystrojem wnętrza. Dzięki temu zwiedzanie rezydencji jest tak ciekawe, a za każdym zakrętem pałacowego korytarza czeka na zwiedzającego kolejna niespodzianka. Sytuacja ta jest jednak także przyczyną poważnych wyzwań stojących przed muzeum. Tworzenie programu precyzyjnej dokumentacji trójwymiarowej wybranych wnętrz jest jedną z prób radzenia sobie z tymi problemami. Skomplikowany układ przestrzenny pałacu utrudnia zapewnienie dostępu zwiedzającym do wszystkich pomieszczeń. Ze względu na ograniczenia konserwatorskie i względy bezpieczeństwa do niektórych pomieszczeń turyści mogą jedynie zajrzeć przez otwarte drzwi, a czasami nawet taka forma udostępnienia nie jest możliwa. Stworzenie i udostępnienie wysokiej jakości dokumentacji cyfrowej zapewnia tym wspaniałym wnętrzom możliwość funkcjonowania w powszechnej świadomości społecznej. Inny problem wynika z faktu, że Pałac budowany jako rezydencja letnia, charakteryzujący się cienkimi murami, funkcjonuje dziś jako całoroczne muzeum w trudnych warunkach polskiego

klimatu. Brak możliwości położenia odpowiednich izolacji termicznych powoduje, że zapewnienie odpowiednich warunków środowiskowych we wnętrzach pałacu w różnych porach roku jest dużym wyzwaniem. Monitorowanie stanu zachowania malarstwa ściennego, drewnianych polichromowanych zdobień i innych elementów wystroju wnętrz jest zadaniem łatwiejszym, gdy posiada się bardzo precyzyjną dokumentację przestrzenną dającą możliwość weryfikacji nawet drobnych zmian jakie zachodzą z biegiem czasu.

Dane w postaci chmur punktów wykorzystane w niniejszej rozprawie pochodzą z czterech pomieszczeń posiadających zbliżone wymiary wynoszące około 4,0 [m] na 4,0 [m] oraz wysokość 5,0 [m]. Gabinet Chiński Króla (rysunek 9a) i Garderoba Króla (rysunek 9b) to dwa pomieszczenia znajdujące się w południowej użytkowanej przez króla części pałacu. Pozostałe dwa pokoje, Antygabinet Królowej (rysunek 9c) oraz Gabinet al fresco (rysunek 9d) położone są po drugiej stronie osi głównej pałacu w części użytkowanej przez królową.



Rysunek 9: Zdjęcia skanowanych pomieszczeń: a) Gabinet Chiński Króla [182], b) Garderoba Króla [183], c) Antygabinet Królowej [184], d) Gabinet al fresco (narożnik północno-zachodni) [185].

Obecna dekoracja Gabinetu Chińskiego Króla (rysunek 10) jest dziełem warsztatu Martina Schnella, nadwornego lakiernika i malarza króla Augusta II Mocnego. Dekoracja ścienna, powstała około 1730 roku, ma postać polichromowanych drewnianych paneli, malowanych laką, a następnie pokrywanych drobnymi kawałkami posrebrzanej miedzi, co nadaje dekoracji charakterystyczny blask. Sufit pokryty jest malowidłem ściennym, które nawiązuje tematycznie i kolorystycznie do dekoracji ściennej, ale charakteryzuje się znacznie mniejszym odblaskiem.



Rysunek 10: Ortoobraz Gabinetu Chińskiego Króla [186].

Garderoba Króla (rysunek 11) posiada dekorację stropu z czasów króla Jana III oraz dekorację ścienną wykonaną po 1730 r. Tutaj również ściany pokryte są drewnianymi boazeriami, w które wkomponowano malowidła wykonane przez grupę artystów saksońskich. Ich tematyka podkreśla związek wnętrza pałacu z otaczającą go przyrodą ogrodów. W dekoracji tego pomieszczenia dominują jasne barwy połączone z dużą ilością złożonych powierzchni.



Rysunek 11: Ortoobraz Garderoby Króla [186].

Antygabinet Królowej jest pomieszczeniem, którego dekoracja jest datowana na rok 1732, kiedy Pałac w Wilanowie był używany przez króla Augusta II Mocnego. Iluzjonistyczne malarstwo plafonowe w tym pomieszczeniu jest dziełem Jules Poisona. Ściany dekorowane są scenami nawiązującymi do mitologii greckiej, opisanymi przez Owid w dziele „Metamorphoses”. Ortoobraz pomieszczenia został przedstawiona na rysunku 12.



Rysunek 12: Ortoobraz Antygabinetu Królowej [186].

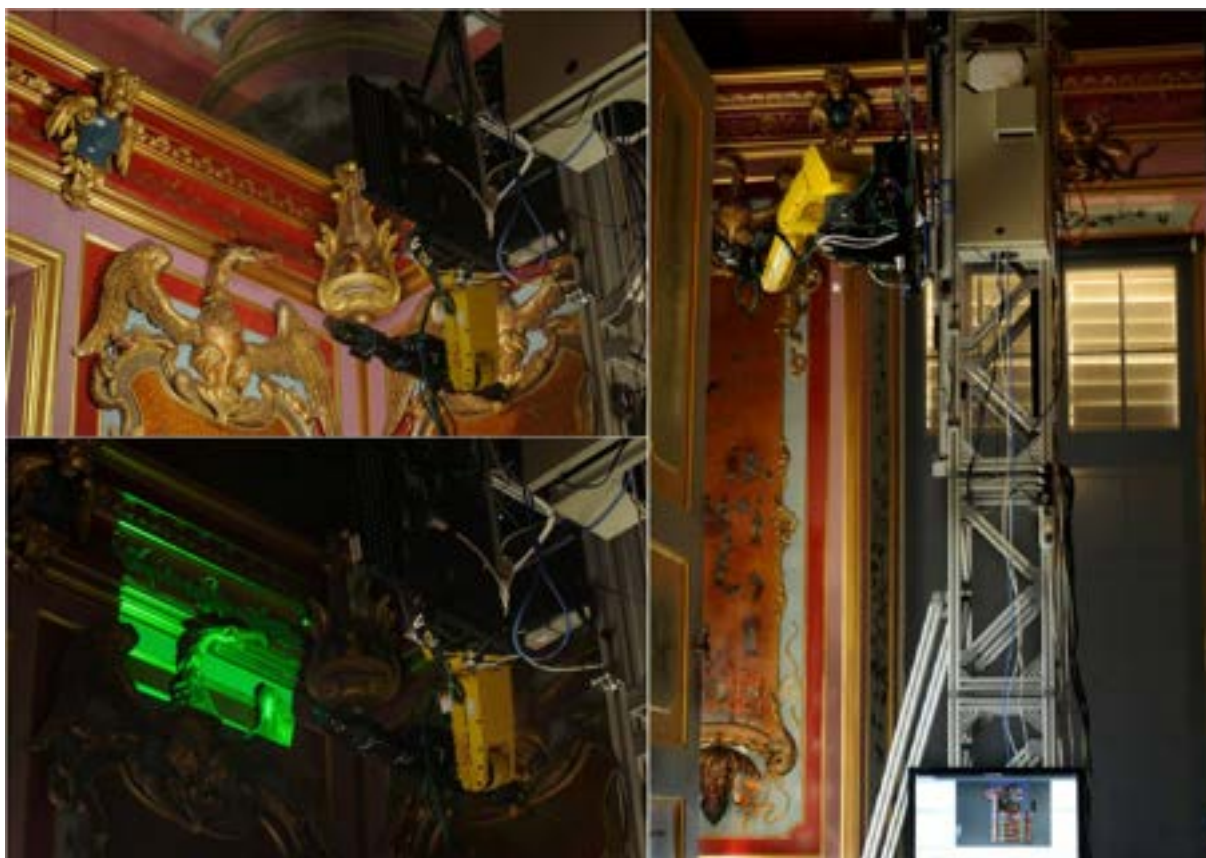
Kolejnym pokojem jest Gabinet Al Fresco (rysunek 13). Jest ozdobiony typową włoską dekoracją namalowaną techniką al fresco przez Jerzego Eleutera Siemiginowskiego. Nie zachowana do dziś dekoracja sklepienia również miała taki sam charakter. Podczas konserwacji przeprowadzonej w 1955 roku odkryto malowidła ściennie, które naśladują zawieszony w narożach trzy gobeliny z figuralnymi przedstawieniami oraz imitują kamienne obramowanie okien i drzwi.



Rysunek 13: Ortoobraz Gabinetu Al Fresco [186].

3.2 System pomiarowy

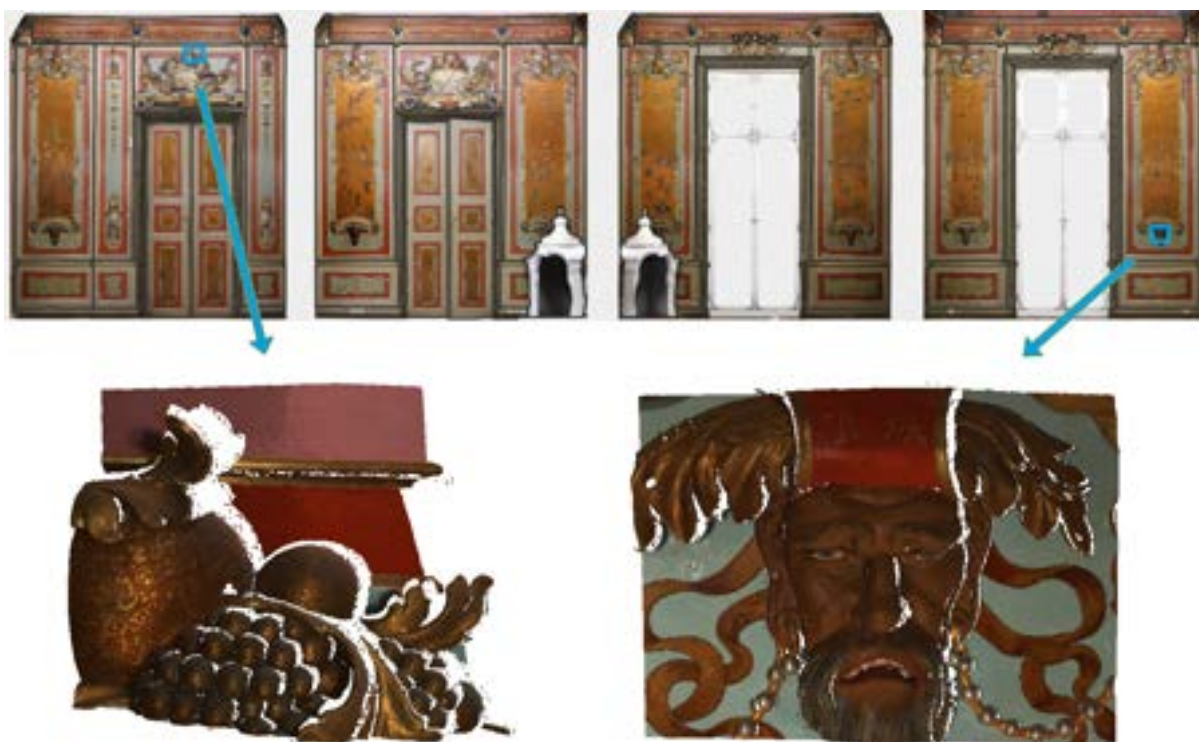
Źródłowe chmury punktów 3D zostały zarejestrowane za pomocą dedykowanego systemu pomiarowego opracowanego w ramach kampanii akwizycji wnętrza [187] (rysunek 14). System ten został opracowany w celu częściowej automatyzacji akwizycji, co pozwoliło na przyspieszone, w stosunku do istniejących metod, uzyskanie wysokiej jakości danych pomiarowych pod względem rozdzielczości, precyzji oraz koloru. System ten składa się ze skanera 3D wykorzystującego światło strukturalne do rejestrowania informacji o kształcie oraz kamery służące do akwizycji barwy. W trakcie akwizycji zbierano sześć obrazów z różnymi kierunkami oświetlenia, co pozwoliło na redukcję szumu pomiarowego związanego z odbiciem światła. Dwa oświetlacze zamontowano na głowicy pomiarowej, natomiast cztery umieszczono na podstawie robota. Sam robot, na którym zamontowano skaner 3D, został wyposażony w pionowy podnośnik o maksymalnej wysokości roboczej wynoszącej 5,5 metra. Cały system oparty był na stabilizowanej platformie, wyposażonej w koła, co znacznie ułatwiało jego przemieszczanie.



Rysunek 14: System pomiarowy zastosowany do wysokorozdzielczej dokumentacji 3D pomieszczeń zabytkowych.

3.3 Dane

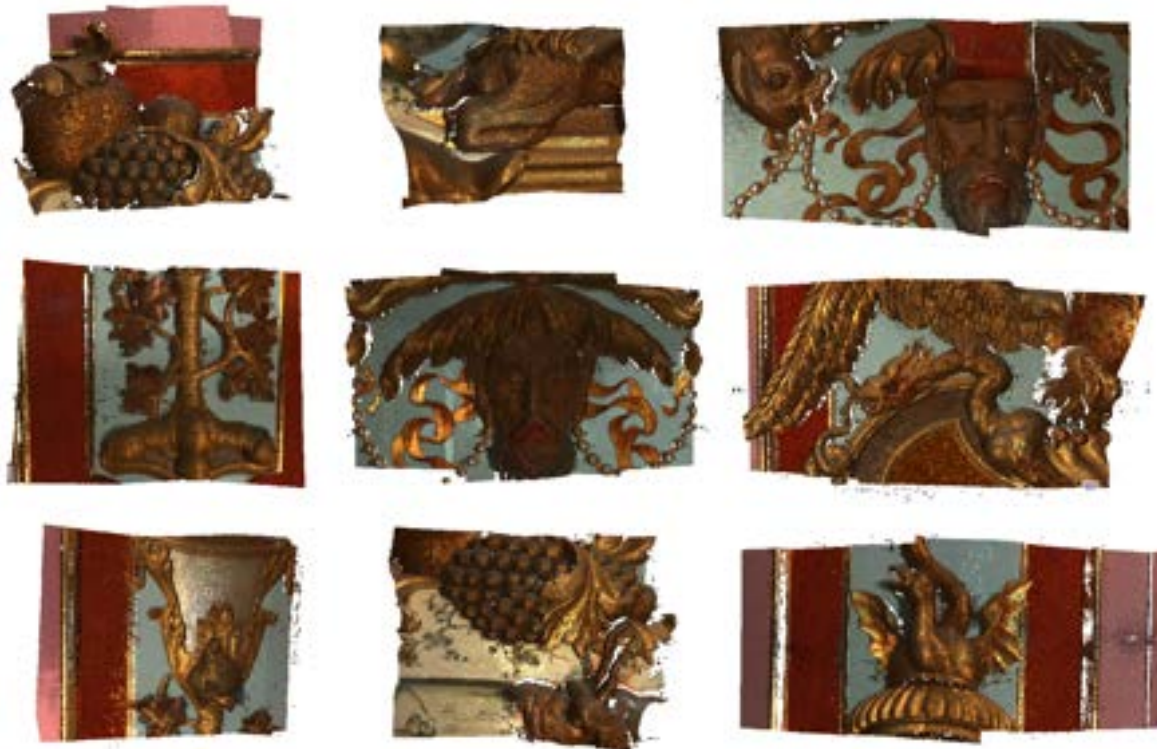
Liczba chmur punktów zarejestrowanych podczas digitalizacji 3D pojedynczego pomieszczenia jest stosunkowo duża (około pięciu tysięcy dla każdego pomieszczenia). Każda chmura punktów zawiera około 7,5 miliona punktów (patrz rysunek 15). Błąd geometryczny chmur punktów jest niższy niż 0,05 [mm], przy średniej odległości między punktami wynoszącej około 0,1 [mm]. Każdy punkt w chmurze jest reprezentowany przez swoje współrzędne trójwymiarowe (x, y, z), wektor normalny (n_x , n_y , n_z) oraz skalibrowane wartości barwy (R, G, B).



Rysunek 15: Chmury punktów w porównaniu do całego pomieszczenia Gabinetu Chińskiego Króla.

Zbiór danych użyty w niniejszej rozprawie w ramach eksperymentu oraz testów jest podzbiorem wszystkich chmur punktów pochodzących z czterech skanowanych pomieszczeń. Dane zostały podzielone na pięć podzbiorów, każdy składający się z 25 par chmur punktów. Różnią się one między sobą poziomem szumu, geometrią, kolorem i teksturą. Pary chmur punktów posiadają różne wielkości części wspólnej - ich wzajemne pokrycie waha się w przedziale od około 20% do 90%.

Pierwszy podzbiór obejmuje chmury punktów o szczególnej geometrii, dokładnie odzwierciedlającej kształt powierzchni. Są one głównie płaskorzeźbami, bądź ich fragmentami zdobiącymi pomieszczenia pałacu. Przykłady tych chmur można zobaczyć na rysunku 16.



Rysunek 16: Dziewięć przykładowych par chmur punktów - podzbiór 1.

Drugi wybrany podzbiór składa się z płaskich chmur punktów, charakteryzujących się różnorodnymi kolorami, głównie przedstawiających obrazy i dzieła sztuki wykonane różnymi technikami. Dziewięć przykładowych chmur punktów pochodzących z drugiego podzbioru można zobaczyć na rysunku 17.



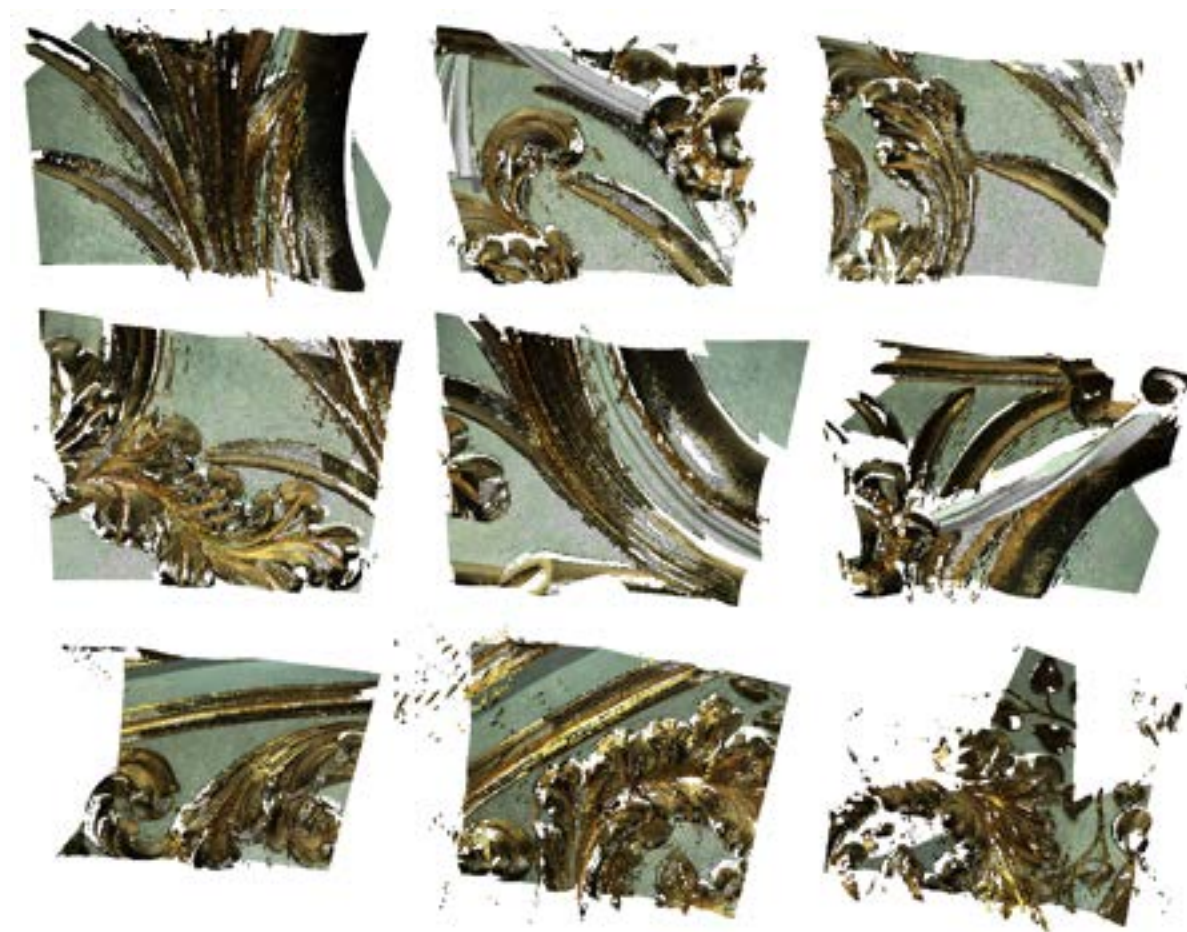
Rysunek 17: Dziewięć przykładowych par chmur punktów - podzbiór 2.

Trzeci podzbiór łączy cechy dwóch poprzednich, prezentując dekoracyjne malowidła na zakrzywionych sufitach (rysunek 18). Poza krzywizną, sufit charakteryzuje się chropowatą powierzchnią oraz pęknięciami.



Rysunek 18: Dziewięć przykładowych par chmur punktów - podzbiór 3.

Czwarty i zarazem najtrudniejszy podzbiór zawiera fragment pomieszczenia zawierający liczne pozłacane i błyszczące dekoracje (rysunek 19 b). Chmury punktów należące do tej podkategorii charakteryzują się wysokim poziomem szumów pomiarowych.



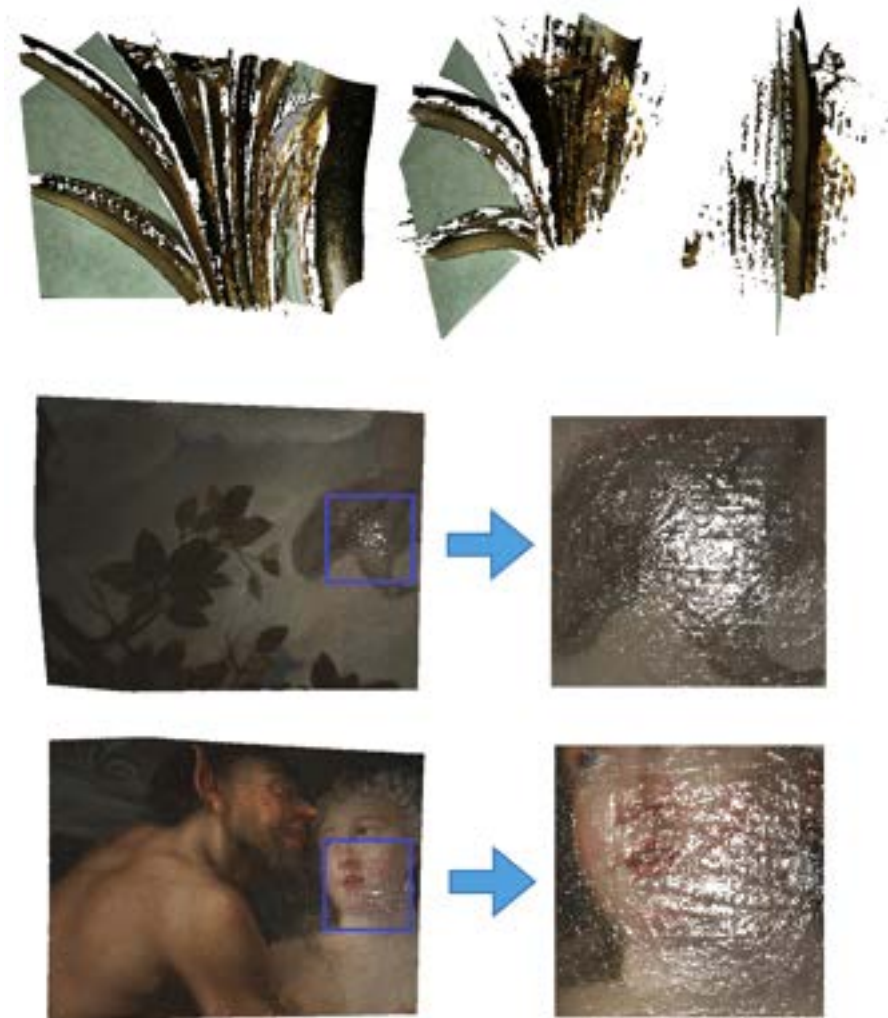
Rysunek 19: Dziewięć przykładowych par chmur punktów - podzbiór 4.

Ostatni, piąty podzbiór składa się chmur punktów, w których odwzorowano płaskie fragmenty ścian. Charakteryzują się one jasną barwą oraz inną teksturą powierzchni w stosunku do chmur punktów pochodzących z innych pomieszczeń. Dodatkowo występują tutaj powtarzające się wzory, które znacząco utrudniają wstępne dopasowanie. Przykłady dziewięciu par chmur punktów znajdują się na rysunku 20.



Rysunek 20: Dziewięć przykładowych par chmur punktów - podzbiór 5.

Wszystkie dane zebrane przez skaner 3D charakteryzują się występowaniem odbić świetlnych na powierzchniach błyszczących. Pomimo użycia sześciu źródeł światła podczas skanowania, uzyskane chmury punktów wciąż zawierają artefakty i szumy spowodowane tym czynnikiem (rysunek 21).



Rysunek 21: Artefakty spowodowane odbiciami na skanowanej powierzchni.

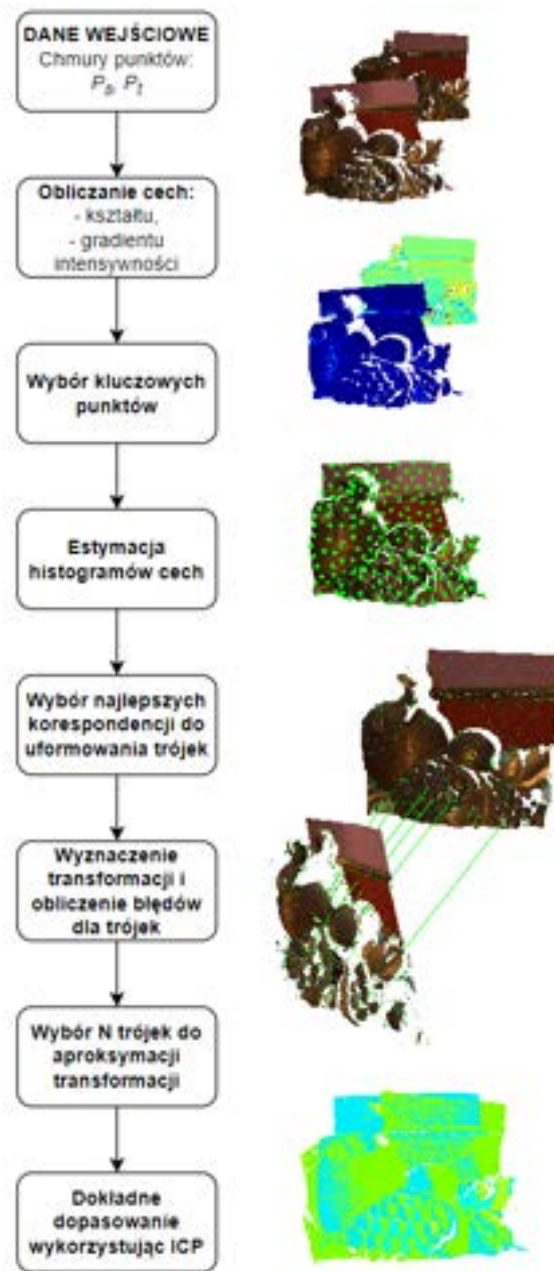
4 Koncepcja

Pierwsze próby automatyzacji i przyspieszenia składania tak dużej ilości chmur punktów pochodzących z projektów dokumentacyjnych w Wilanowie zostały podjęte już wcześniej [188]. Celem było zmniejszenie lub całkowite wyeliminowanie problemów występujących przy procesie łączenia chmur punktów napotkanych podczas cyfrowych rekonstrukcji pomieszczeń realizowanych w Pałacu w Wilanowie. Poprzez odpowiednie przygotowanie danych, uproszczenie, wyznaczenie odpowiednich warstw danych oraz korekcję barwy znacząco skrócono czas potrzebny na poprawne połączenie pojedynczej pary chmur punktów. Główne założenia metody opierają się na wykorzystaniu klasycznego algorytmu ICP. Istotną rolę odgrywa odpowiednia organizacja danych w pamięci komputera. Dane dzielone były na mniejsze fragmenty, złożone z kilkadziesiąt chmur. Każdy z tych fragmentów był oddzielnie przetwarzany - operator wybierał jedną chmurę jako początkową, a każda kolejna była automatycznie wybierana ze zbioru na podstawie wielkości części wspólnej pomiędzy nimi. Dla danych geometrycznych wykorzystano klasyczny algorytm ICP oparty na odległości, a dla danych planarnych z barwą warunkiem wyboru najbliższego punktu była obliczona cecha bazująca na kolorze oraz różnicy Gaussa [189].

Aby cały proces mógł przebiegać prawidłowo, musiał być spełniony jeden warunek: cały system do skanowania musiał być poprawnie i precyzyjnie skalibrowany, aby dostarczyć wstępne, względne transformacje 3D zebranych chmur punktów. Pozwala to już na początku zgrubnie określić położenie każdej z chmur punktów w modelu i uniknąć czasochłonnego wstępnego dopasowywania. Niestety warunek ten nie zawsze był spełniony. Często wstępne dopasowanie musiało być przeprowadzone manualnie, co znacząco wpływało na czas potrzebny na przetworzenie całego modelu.

Aby rozwiązać ten problem, w niniejszej rozprawie zaproponowano proces orientacji par chmur punktów o nazwie FAMFR (Fast Adaptive Multimodal Feature Registration [190]) polegający na automatycznym wyszukiwaniu przybliżonych wzajemnych transformacji 3D pomiędzy poszczególnymi chmurami punktów. Do uzyskania finalnego modelu zaimplementowano kilka metod złożonych w jeden zautomatyzowany proces. Poglądowy schemat zaproponowanej koncepcji znajduje się na rysunku 22.

Zaproponowana koncepcja opiera się na dwóch cechach wyznaczonych dla każdego z punktów w dopasowywanych chmurach punktów. Na podstawie wyznaczonych cech, jednej bazującej na geometrii oraz drugiej na gradientach intensywności, wybierane są losowo



Rysunek 22: Schemat zaproponowanej koncepcji FAMFR.

próbkowane punkty charakterystyczne. Dla każdego z tych punktów wyznaczane są histogramy cech (geometrycznych oraz gradientów) na podstawie lokalnego otoczenia. Następnie wybierany jest podzbiór najlepszych dopasowań pomiędzy wyznaczonymi histogramami cech oraz tworzony jest zbiór potencjalnych "trójek" par punktów. Aby poprawnie wyznaczyć transformację 3D pomiędzy dwoma chmurami punktów niezbędne jest znalezienie co najmniej trzech odpowiadających sobie par punktów. Taki podzbiór jest poddany analizie zgodności przestrzennej, co skutkuje wstępnym odfiltrowaniem błędnych

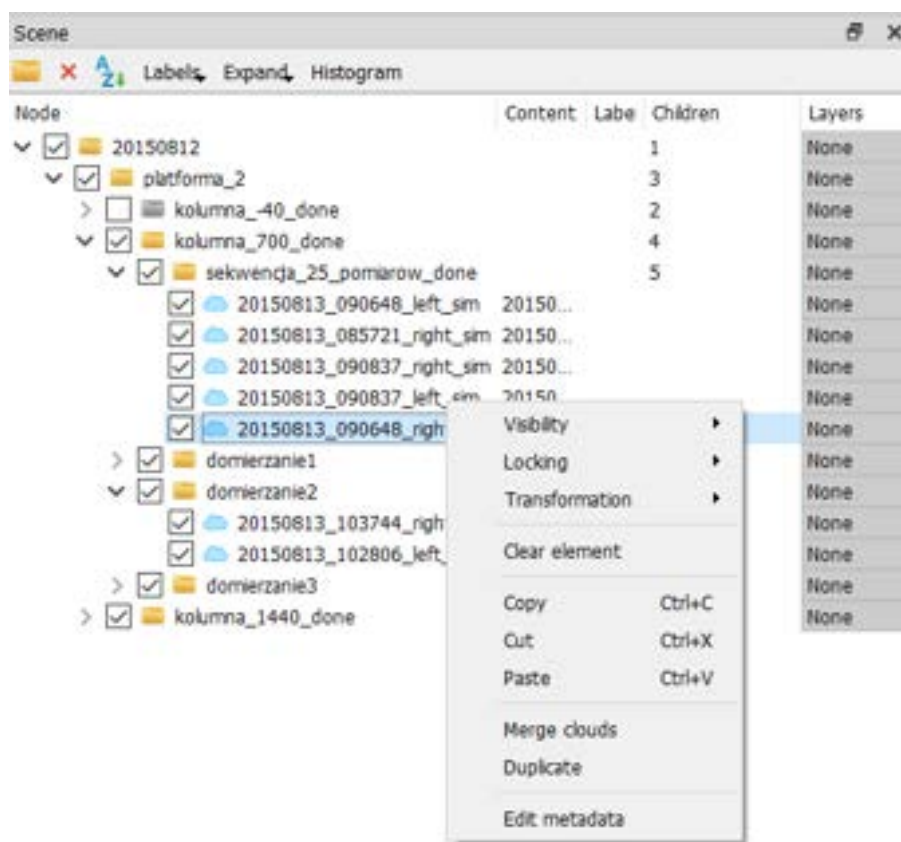
trójek. Z takiego podzbioru przeprowadzana jest analiza błędu dopasowania dla każdej transformacji i wybierane są te korespondencje, które dają najlepsze wyniki. Końcowa transformacja 3D jest otrzymywana poprzez wykorzystanie klasycznego algorytmu ICP. Dodatkowo, zdecydowano się uzależnić parametry wejściowe algorytmu od średniej odległości między punktami D_{avg} przetwarzanych chmur punktów. Pozwala to na uogólnienie opracowanej koncepcji i zwolnienie użytkownika z doboru parametrów algorytmu.

5 Implementacja

Metoda FAMFR została zaimplementowana oraz przetestowana na komputerze z parametrami: 64 GB pamięci RAM, procesor Intel Core I7-8850H 2,60 GHz oraz karta graficzna NVIDIA GeForce Quadro P1000.

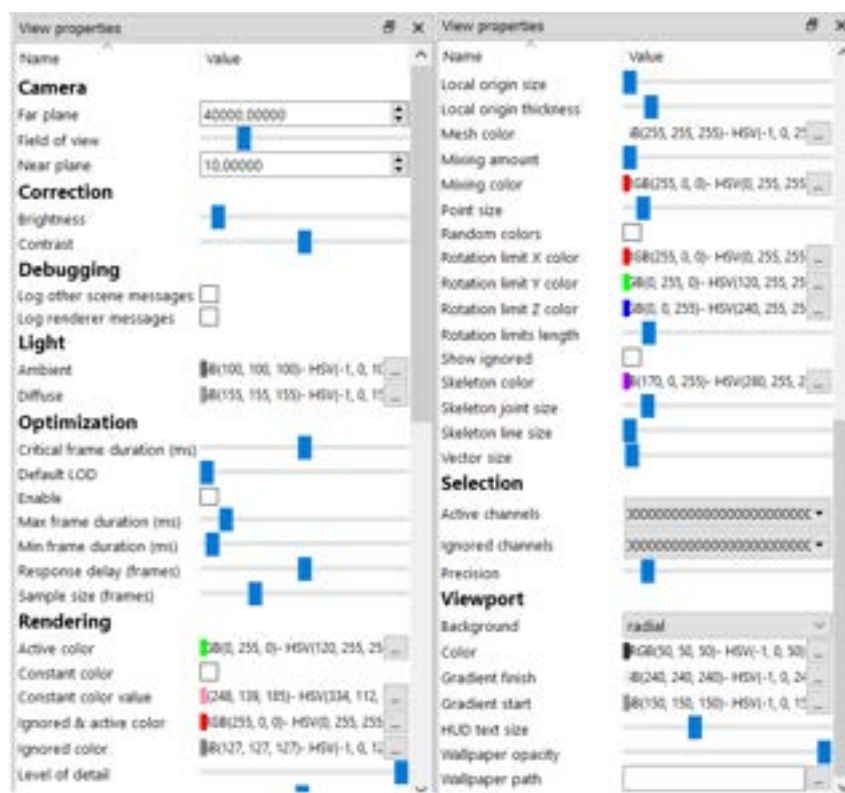
5.1 Środowisko implementacyjne

W niniejszej rozprawie w ramach eksperymentów wykorzystano oprogramowanie FRAMES napisane w języku C++. FRAMES (Framework and Robust Algorithms for Models of Extreme Size) opracowany na Wydziale Mechatroniki Politechniki Warszawskiej, jest narzędziem do przetwarzania i wizualizacji dużych zbiorów danych 3D. Zdolność do zarządzania i przetwarzania dużych zbiorów danych jest jednym z kluczowych atutów oprogramowania FRAMES. Dane w nim są udostępniane z poziomu interfejsu użytkownika wykorzystując strukturę opartą na grafach (rysunek 23), co znacznie ułatwia efektywne manipulowanie dużymi zbiorami i strukturami danych [187].



Rysunek 23: Interfejs FRAMES do obsługi danych.

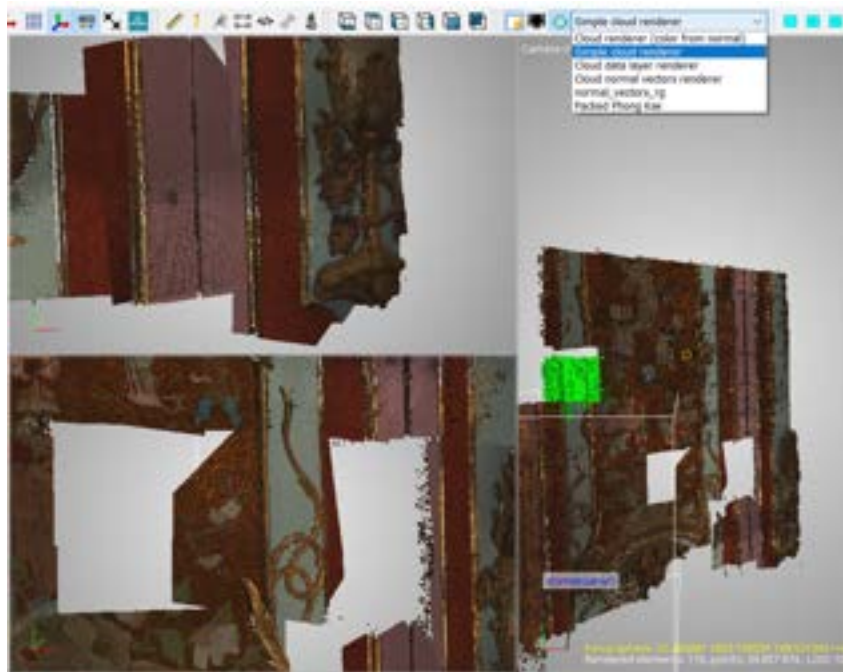
FRAMES zapewnia uproszczony interfejs użytkownika oraz łatwe wizualizowanie wyników dzięki dostępności parametrów odpowiadających za wizualizację, optymalizację oraz wirtualną kamerę. Dodatkowo, efektywna wizualizacja dużych zbiorów danych 3D jest możliwa dzięki dynamicznemu mechanizmowi LoD (Level of Detail), który jest dostępny w tym oprogramowaniu. W przypadku zbyt długiego renderowania aktualnie widzianej sceny, FRAMES automatycznie obniża jakość wyświetlanej klatki, zmniejszając o połowę liczbę widocznych punktów 3D. To przyjazne dla użytkownika podejście usprawnia przepływ pracy i ułatwia interakcję oraz interpretację przez użytkownika analizowanych chmur punktów. Na rysunku 24 zestawiono cały zbiór dostępnych parametrów wizualizacji.



Rysunek 24: Interfejs FRAMES do modyfikacji parametrów wizualizacji.

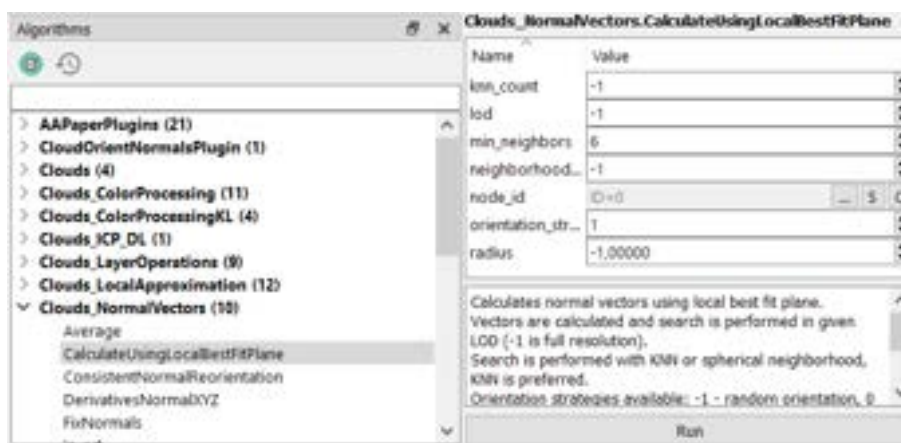
Poza tym co zostało opisane, w ramach wizualizacji FRAMES umożliwia również podział głównego widoku na mniejsze instancję, szybkie przełączanie widoków czy zmianę dostępnego mechanizmu renderującego (rysunek 25).

FRAMES oferuje szeroki zakres zaawansowanych algorytmów między innymi do filtracji szumów, segmentacji chmur punktów, wyznaczania punktów charakterystycznych oraz ich opisu za pomocą deskryptorów, dopasowywania chmur punktów i wiele innych. Interfejs umożliwia również łatwe wprowadzanie i modyfikowanie parametrów wraz z dostępną



Rysunek 25: Interfejs FRAMES do obsługi sceny.

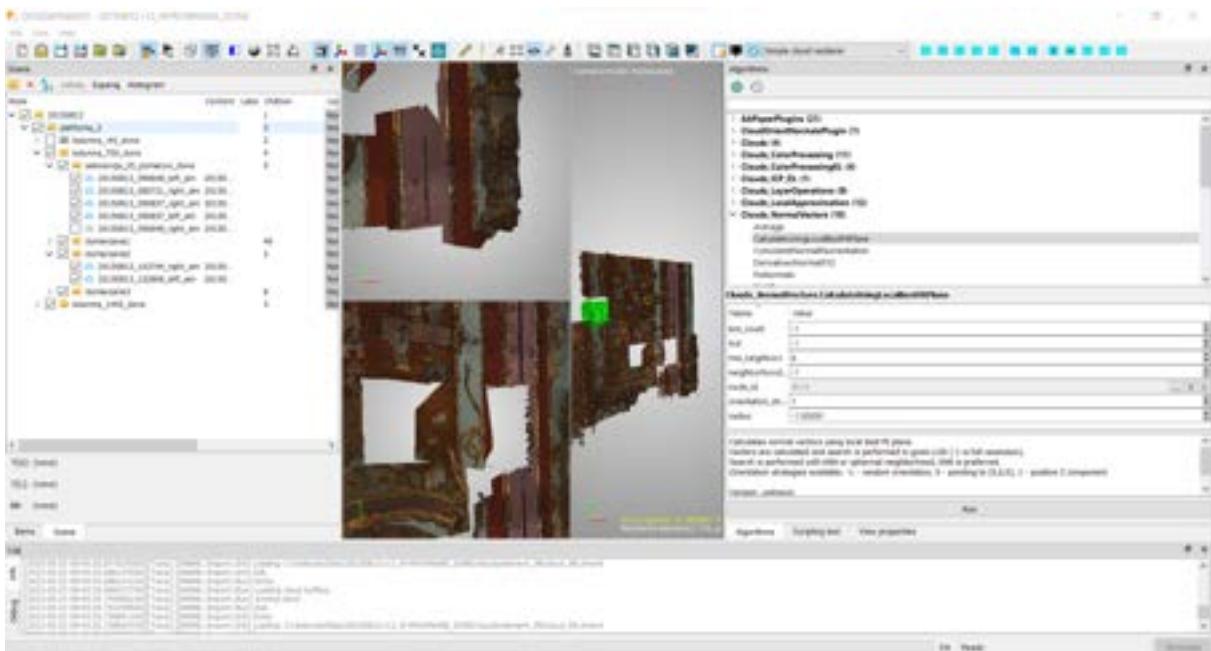
historią uruchamianych algorytmów. Możliwość rozszerzania funkcjonalności środowiska FRAMES zapewnia zaimplementowany system wtyczek, który upraszcza rozwój i implementację niestandardowych algorytmów do przetwarzania danych 3D. Wbudowane algorytmy, w połączeniu z możliwością rozszerzenia funkcjonalności oprogramowania za pomocą własnych algorytmów, dostarczają kompleksowego zestawu narzędzi do przetwarzania i analizy chmur punktów 3D. Ta elastyczność pozwala dostosować oprogramowanie do swoich konkretnych potrzeb. Na rysunku 26 zaprezentowano interfejs wtyczek wraz z pojedynczym algorytmem i jego parametrami.



Rysunek 26: Interfejs FRAMES do obsługi algorytmów.

Dodatkowo, pod względem obsługi danych, FRAMES również pozwala na przetwarzanie pomiarów 4D (sekwencje skanów 3D z różnych chwil czasowych) [191]. Ze względu na ograniczoną pamięć RAM, oprogramowanie ładuje dane partiami lub klatkami. W pamięci przechowywane są tylko przetwarzane aktualnie klatki oraz kilka klatek poprzedzających oraz kolejnych. Takie podejście zapewnia płynne nawigowanie i wizualizację chmur punktów, nawet przy pracy na dużych zbiorach danych.

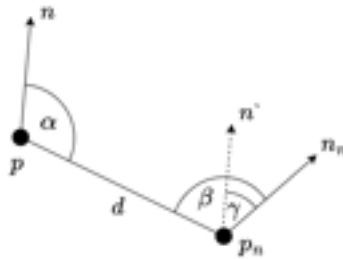
Podsumowując, FRAMES jest wszechstronnym narzędziem do szybkiego przetwarzania danych 3D/4D. Rozbudowany system wtyczek wraz z dostępnymi algorytmami pozwala na bardzo szybkie rozpoczęcie pracy, a możliwość pisania skryptów na natychmiastowe prototypownie dedykowanych rozwiązań. Wysoka wydajność oraz łatwy do opanowania interfejs użytkownika pozwala na analizę złożonych zestawów danych, dzięki czemu doskonale nadaje się do zastosowań związanych z dziedzictwem kulturowym, takich jak łączenie chmur punktów. Całe okno FRAMES można zobaczyć na rysunku 27.



Rysunek 27: Główne okno oprogramowania FRAMES.

5.2 Przetwarzanie wstępne

W celu przyspieszenia procesu łączenia chmur punktów oraz zmniejszenia zapotrzebowania na wymaganą pamięć komputera, dane zostały jednorodnie uproszczone parametrem N_{sim} . Parametr ten wyznacza stosunek liczby punktów wejściowej chmury do liczby punktów po uproszczeniu, a jego wartość została dobrana w taki sposób, aby nie wpływała negatywnie na precyzję wymaganą przy łączeniu dwóch chmur punktów. Po uproszczeniu wyznaczany jest parametr D_{avg} reprezentujący średnią odległość między punktami w przetwarzanych chmurach punktów, który następnie jest wykorzystywany do parametryzowania kolejnych etapów metody FAMFR. Kolejnym krokiem jest wyznaczenie wektorów cech dla obu chmur: chmury stabilnej/referencyjnej (P_s) oraz chmury dopasowywanej (P_t). Pierwszy z nich jest obliczany następująco: dla każdego z punktów p tworzona jest sfera o promieniu $R_n \cdot D_{avg}$ obejmująca lokalne sąsiedztwo. Z tak wyznaczonej sfery generowany jest wektor cech opierający się na odległościach pomiędzy zadany punkt p i sąsiadującymi punktami p_n oraz na zależnościach kątowych pomiędzy wektorami normalnymi skojarzonymi z punktami: n oraz n_n . Dokładne relacje są zaprezentowane na rysunku 28 i obliczane według wzoru 1.



Rysunek 28: Relacje kątowe opisujące cechę geometryczną par punktów.

$$V_s = (\angle(n, d), \angle(n_n, d), \angle(n, n_n)) \quad (1)$$

Drugą cechą wykorzystaną do lokalnego opisu chmury punktów jest gradient intensywności. Gradient intensywności I_g w zadany punkcie p jest wektorem ortogonalnym do wektora normalnego, skierowanym w kierunku maksymalnego wzrostu lokalnej intensywności. Deskryptor ten również budowany jest na podstawie lokalnego otoczenia. Pierwszym etapem jest estymacja wartości jasności RGB (ang. luminosity) dla każdego z punktów p_n w lokalnym otoczeniu zgodnie ze wzorem 2.

$$L = 0.299 \cdot red + 0.587 \cdot green + 0.114 \cdot blue \quad (2)$$

Współczynniki wykorzystane we wzorze 2 (0.299, 0.587, 0.114, ang. luma coefficients) bazują na funkcji wyznaczającej względną jasność [192] opartej na matematycznym modelu percepcji barw przez ludzkie oko. Na podstawie wyznaczonych wartości w lokalnym otoczeniu obliczana jest średnia wartość jasności L_{avg} oraz uśrednione współrzędne punktów p_{avg} . Następnie dla każdego punktu obliczana jest lokalna różnica w jasności L_d pomiędzy wartością średnią L_{avg} oraz L_n według wzoru 3.

$$L_{avg} = \frac{\sum L_n}{n}$$

$$L_d = L - L_{avg} \quad (3)$$

Również współrzędne każdego punktu w lokalnym otoczeniu p_n są modyfikowane średnią wartością p_{avg} według wzoru 4.

$$p_n = p_n - p_{avg} \quad (4)$$

W następnym kroku budowane są macierz A oraz wektor b korzystając ze współrzędnych 3D lokalnych punktów według wzoru 5.

$$A = \begin{bmatrix} \sum p_{nx} \cdot p_{nx} & \sum p_{nx} \cdot p_{ny} & \sum p_{nx} \cdot p_{nz} \\ \sum p_{nx} \cdot p_{ny} & \sum p_{ny} \cdot p_{ny} & \sum p_{ny} \cdot p_{nz} \\ \sum p_{nx} \cdot p_{nz} & \sum p_{ny} \cdot p_{nz} & \sum p_{nz} \cdot p_{nz} \end{bmatrix} \quad (5)$$

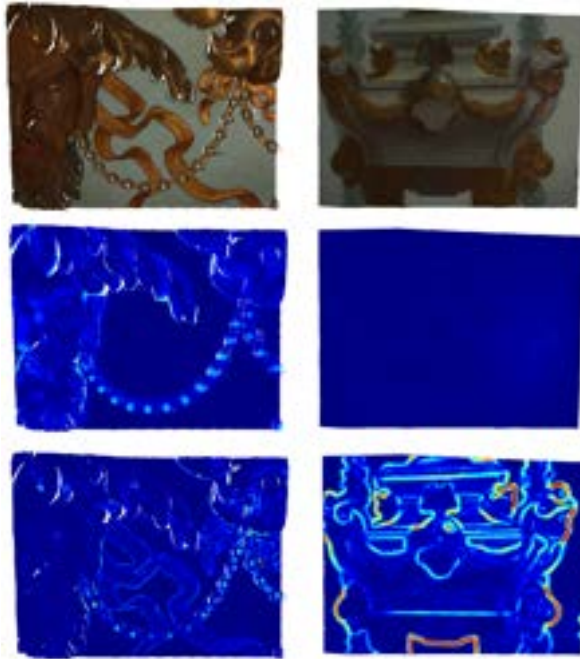
$$b = \begin{bmatrix} \sum p_{nx} \cdot L_d & \sum p_{ny} \cdot L_d & \sum p_{nz} \cdot L_d \end{bmatrix}$$

$$x = A^{-1} \cdot b$$

Tak utworzona macierz poddawana jest dekompozycji QR metodą Householdera [193], w wyniku której otrzymujemy wektor x . Ostateczny wektor intensywności gradientów $I_g(L_x, L_y, L_z)$ jest utworzony z trzech wartości na podstawie wzoru 6:

$$I_g = (Identity - n \cdot n^T) \cdot x \quad (6)$$

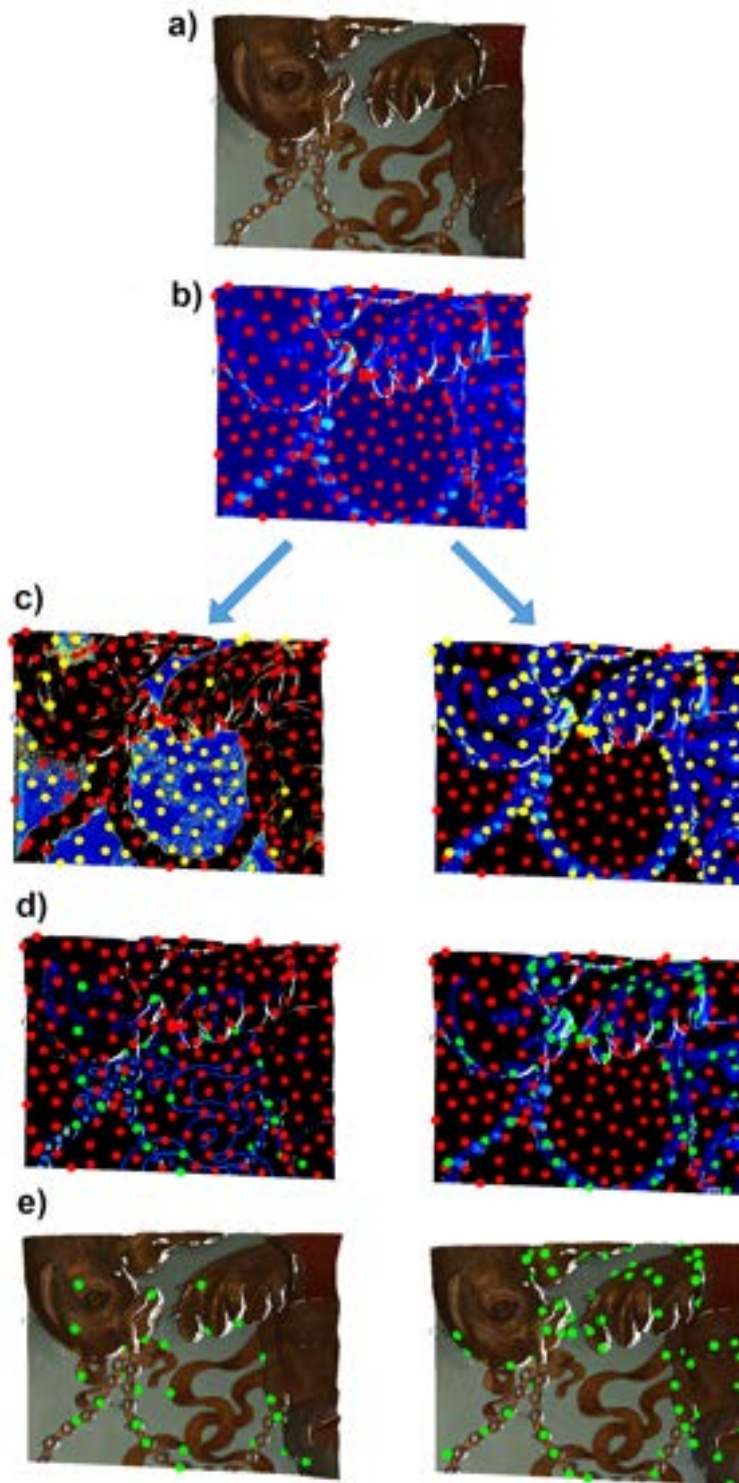
Gdzie *Identity* reprezentuje macierz jednostkową, a n wektor normalny. Wizualizację obu tych cech za pomocą wielkości utworzonych wektorów zaprezentowano na rysunku 29.



Rysunek 29: Wyznaczone wielkości cech: na górze wejściowe chmury punktów, w środku wielkość wektora reprezentującego kształt $|V_s|$, na dole wielkość lokalnego gradientu $|V_g|$.

5.3 Wyznaczanie punktów charakterystycznych

Dwie wyznaczone cechy w postaci wektorów odgrywają również kluczową rolę w identyfikacji potencjalnych punktów charakterystycznych. Na początku wyznaczane są dwa równomiernie próbkowane podzbiory punktów (P_{ss} oraz P_{ts}) z chmury referencyjnej P_s oraz dopasowywanej P_t z zachowaniem odległości pomiędzy wybranymi punktami równej $K_d \cdot D_{avg}$. Następnie punkty są dzielone na dwie kategorie, bazując na wyznaczonych cechach. Pierwsza grupa odpowiada za dopasowywanie chmur punktów wykorzystując informację o kształcie V_s , druga wykorzystuje do tego samego celu wyznaczone gradienty V_g . Liczba punktów charakterystycznych jest duża, dlatego wybrane podzbiory przechodzą proces filtracji poprzez odrzucenie punktów, których wielkość danego wektora cech jest mniejsza od zadanego progu T_h . T_h jest procentową wartością w stosunku do maksymalnej wartości cechy dla całej chmury punktów ($max(|V_s|)$ albo $max(|V_g|)$). Cały proces został przedstawiony na rysunku 30.

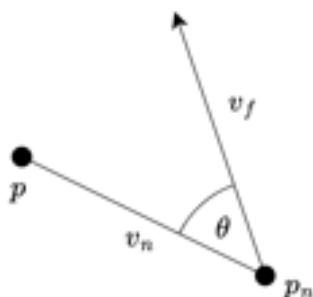


Rysunek 30: Wyznaczanie punktów charakterystycznych na podstawie cech geometrycznych (po prawej) oraz gradientów (po lewej). a) wejściowe chmury punktów, b) równomiernie wybrane punkty charakterystyczne, c) podział punktów na dwie kategorie, d) filtracja punktów charakterystycznych (T_h), e) Finalnie wybrane punkty do procesu łączenia chmur punktów.

5.4 Histogramy cech

Dla każdego punktu ze zbiorów P_{ss} oraz P_{ts} wyznaczany jest histogram wartości kątowych (θ). Histogram generowany jest z lokalnego otoczenia utworzonego przez sferę o zadanym promieniu różnym dla kształtu H_s oraz gradientów H_g . Wartości promieniu są sparametryzowane średnią odległością pomiędzy punktami D_{avg} . Wartości kątowe są wyznaczane jako kąt odchylenia wektora cech (V_s albo V_g) w każdym z punktów w lokalnym otoczeniu p_n w stosunku do wektora stworzonego z danego punktu charakterystycznego p oraz punktu sąsiadującego p_n . Procedura wyznaczania histogramów została zaprezentowana na rysunku 31 oraz opisana wzorem 7.

$$v_n = p_n - p$$
$$\theta = v_n \cdot \cos\left(\frac{v_n \cdot V}{|v_n||V|}\right) \quad (7)$$



Rysunek 31: Kąt θ wyznaczany dla każdego punktu w otoczeniu lokalnym.

Histogram jest podzielony na ustaloną liczbę komórek parametrem H_b , do których przypisywane są wyznaczone wartości kątowe. Jeżeli obliczony kąt dla danego punktu p_n znajduje się w przedziale danej komórki, wtedy jej wartość zwiększana jest o wielkość wektora podzielona przez dystans od punktu charakterystycznego p ($d = |p_n - p|$). Po wyznaczeniu wartości kątowych dla wszystkich punktów w lokalnym otoczeniu oraz przypisaniu ich do odpowiedniej komórki, histogram normalizowany jest liczbą punktów sąsiadujących. Przykład histogramów dla konkretnych punktów zaprezentowano na rysunku 32.



Rysunek 32: Przykłady histogramów wyznaczonych dla czterech różnych punktów charakterystycznych.

5.5 Wyznaczanie korespondencji

Wyznaczone histogramy służą do znalezienia korespondencji pomiędzy zbiorami punktów P_{ss} oraz P_{ts} . Do oceny podobieństwa pomiędzy poszczególnymi cechami wykorzystano współczynnik S , inspirowany podobieństwem cosinusowym [194]. Jest on wyznaczany za pomocą wzoru 8, gdzie $f_{p_s}(i)$ i $f_{p_t}(i)$ są wartościami poszczególnych komórek histogramów wyznaczonych dla stabilnych/referencyjnych p_s i transformowanych p_t punktów charakterystycznych. Im niższy współczynnik podobieństwa, tym lepsze dopasowanie pomiędzy punktami. Zbiór wszystkich wyznaczonych w ten sposób korespondencji jest filtrowany poprzez wybór określonej liczby (N_p) dopasowanych par punktów.

$$S = \sqrt{\frac{\sum_{i=1}^{H_b} (f_{p_s}(i) - f_{p_t}(i))^2}{\sum_{i=1}^{H_b} (f_{p_s}(i) + f_{p_t}(i))^2}} \quad (8)$$

Do poprawnej estymacji transformacji wymagane są minimum trzy różne, niewspółliniowe punkty. Z wcześniej odfiltrowanych korespondencji formowane są wszystkie potencjalne trójki. Są one następnie poddawane analizie zgodności przestrzennej poprzez podobieństwo trójkątów (T_s). Jest to wartość opisująca geometryczny koncept, który zakłada, że trzy punkty na obu chmurach muszą mieć podobne wzajemne relacje oraz stworzone z nich trójkąty mają podobny kształt i wymiary. Sprawdzenie odbywa się na podstawie analizy długości boków trójkątów, tj. odpowiadające sobie boki trójkąta stworzone z trójek na obu chmurach muszą mieć podobne wymiary. Zakładając dwa trójkąty złożone z trzech punktów A, B, C dla chmury stabilnej oraz

A', B', C' dla chmury transformowanej muszą być spełnione warunki (9).

$$\begin{aligned}T_s &> |||AB|| - ||A'B'||| \\T_s &> |||BC|| - ||B'C'||| \\T_s &> |||CA|| - ||C'A'|||\end{aligned}\tag{9}$$

Dodatkową miarą opisującą trzy pary korespondencji jest połączona wartość podobieństwa C_s (10).

$$C_s = S_{P_1} + S_{P_2} + S_{P_3}\tag{10}$$

gdzie S_{P_i} są wartościami współczynnika podobieństwa wektorów cech każdej pary należącej do trójki. Obie te wartości są wykorzystywane do odrzucenia błędnych trójek, co przekłada się na zdecydowane ograniczenie ilości błędnych korespondencji. Tylko określona parametrem N_T liczba trójek o najmniejszej mierze podobieństwa C_s jest wybierana do dalszych etapów procesu.

Dla każdej trójki ze zbioru odfiltrowanych par korespondencji wyznacza się transformację 3D, którą następnie wykorzystuje się do estymacji błędu dopasowania. Jakość dopasowania dwóch chmur punktów jest oparta na dwóch miarach: liczbie poprawnych dopasowań punktów kontrolnych oraz podobieństwie cech w zadanych punktach. Punktami kontrolnymi są wcześniej znalezione punkty charakterystyczne. Pierwsza miara pozwala wyeliminować błędne przekształcenia posiadające duże podobieństwo cech, ale bardzo mało punktów kontrolnych na części wspólnej obu chmur.

W następnym kroku wyznacza się miarę podobieństwa pomiędzy punktami kontrolnymi na chmurze transformowanej, a lokalnymi odpowiednikami w chmurze referencyjnej. Na tej podstawie wybiera się korespondencje, które posłużyły do obliczenia transformacji 3D wykorzystując algorytm Umeyama [195].

Ostatnim krokiem jest wyznaczenie końcowej transformacji 3D pomiędzy chmurami, który realizowany jest poprzez wykorzystanie algorytmu Iterative Closest Point (ICP). W metodzie łączenia opartej na kształcie, w każdej iteracji wybierany jest najbliższy punkt na podstawie odległości, natomiast w metodzie opartej na gradiencie, opisany współczynnik podobieństwa odpowiada za wybór punktu.

5.6 Parametry

Proces FAMFR jest sterowany stosunkowo małą liczbą parametrów wejściowych zebranych w tabeli 7. Są one odpowiedzialne za ogólną wydajność metody, a ich wartości powinny być konfigurowane w oparciu o specyfikę danych wejściowych. Zalecanym podejściem do określenia ich wartości jest eksperymentalna ocena na małym, reprezentatywnym podzbiore chmur punktów danego obiektu. Parametry zostały uzależnione od średniej odległości między punktami, aby uczynić metodę przyjazną dla użytkownika oraz uogólnić metodę do wykorzystania w przyszłości na różnych zbiorach danych. Znacznie ułatwia to korzystanie z metody oraz zwiększa jej użyteczność.

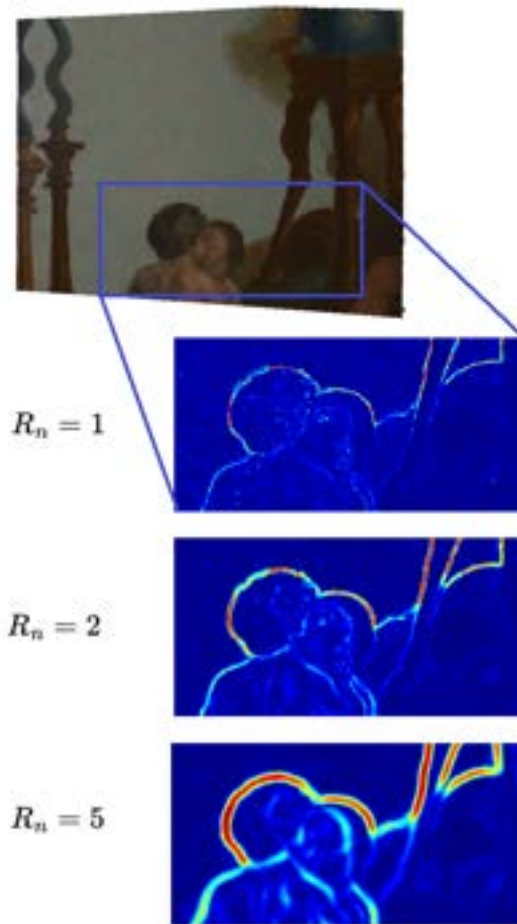
Parametr	Opis
H_b	Liczba komórek histogramu.
N_{sim}	Współczynnik uproszczenia.
R_n	Promień opisujący sferę lokalnego otoczenia do obliczania wektorów cech dla kształtu oraz gradientów.
K_d	Minimalny dystans pomiędzy wybranymi punktami charakterystycznymi.
T_h	Wartość progowa wykorzystywana do filtrowania punktów charakterystycznych.
H_s/H_g	Promień opisujący sferę lokalnego otoczenia do wyznaczania histogramów dla kształtu oraz gradientów.
N_p	Liczba wybranych par punktów do tworzenia trójek korespondencji.
N_T	Liczba wybranych trójek do estymacji transformacji.

Tabela 7: Parametry wejściowe metody FAMFR.

Pierwszy parametr reprezentuje liczbę komórek histogramu. Wartość tego parametru ma najmniejszy wpływ na cały proces. Jego wartość została ustalona eksperymentalnie i wynosi 18.

Kolejnym parametrem jest współczynnik uproszczenia (N_{sim}), który wiąże się z wielkością danych wejściowych. Dla gęstych chmur punktów posiadających dużą liczbę punktów, wartość parametru powinna być stosunkowo wysoka, aby uniknąć zbyt długiego czasu przetwarzania danych. Z drugiej strony, dla chmur z niewielką liczbą punktów wartość parametru nie może być zbyt duża, gdyż może to doprowadzić do utraty opisu kluczowych fragmentów chmury, co by negatywnie odbiło się na liczbie poprawnych korespondencji. Wartość parametru ustalono na 25. W rezultacie, wynikowe chmury punktów posiadają w przybliżeniu około 300 000 punktów oraz średnią odległość pomiędzy punktami równą $D_{avg} = 0.7$ [mm].

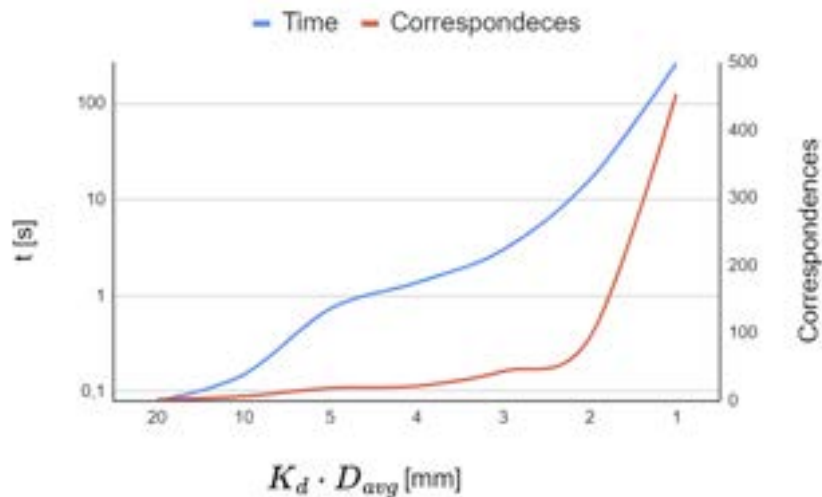
Parametr R_n definiuje długość promienia sfery wyznaczającej lokalne sąsiedztwo wokół punktów. Sąsiedztwo jest wykorzystywane do obliczania wektorów cech kształtu oraz gradientów. Wartość parametru wpływa na precyzję wyliczonych cech. Zmniejszanie jej wartości pozwala uchwycić bardzo drobne detale, a zwiększanie pozwala efektywnie uchwycić powtarzalne punkty charakterystyczne. Przykład wpływu parametru na wynikowe wartości cechy opisującej gradient intensywności można zaobserwować na rysunku 33. Dla danych wykorzystanych w rozprawie wartość została ustalona na $R_n = 7, 7 \cdot D_{avg} \approx 5$ [mm].



Rysunek 33: Wizualizacja danych w zależności od wartości parametru R_n .

Poprawnie dopasowane pary punktów, korespondencje, są kluczowym aspektem całego procesu. To na ich podstawie wyznaczana jest wstępna transformacja. Pierwszym parametrem wpływającym na szukanie korespondencji jest K_d . Odpowiada on za wybór punktów kluczowych, a jego miarą jest minimalny dystans pomiędzy wybranymi punktami. Gdy wartość tego parametru jest zbyt niska, ilość wybranych punktów jest wysoka, co znacząco zwiększa czas potrzebny na wyznaczenie histogramów, porównanie ich oraz odfiltrowanie błędnych dopasowań. Dla danych analizowanych w pracy, znalezienie wstępnych dopasowań

zajmuje około 266 sekund dla wartości $K_d = 2$, jednakże w tym przypadku ilość poprawnych korespondencji jest duża i wynosi nawet 455. Relacje pomiędzy wartością tego parametru, a czasem i liczbą poprawnych dopasowań zostały zaprezentowane na rysunku 34. Podczas eksperymentu wartość parametru ustalono na $K_d = 7,7 \cdot D_{avg} \approx 5$ [mm].

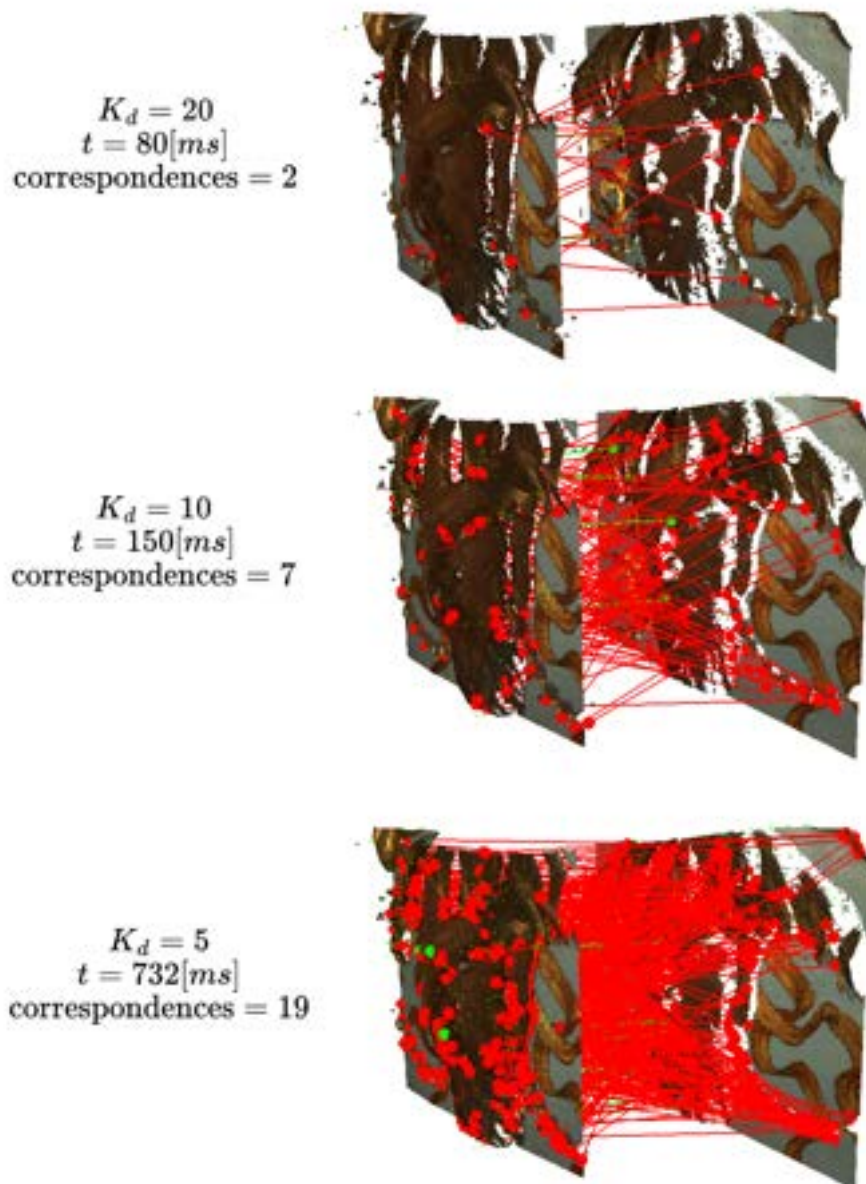


Rysunek 34: Relacja pomiędzy wartością parametru K_d , a czasem oraz liczbą znalezionych korespondencji. Oś czasu przedstawiono w logarytmicznej skali.

Z drugiej jednak strony, zbyt duża wartość tego parametru skutkuje wyborem punktów, które nie posiadają wystarczająco unikatowej cechy do poprawnego znalezienia korespondencji. W tym przypadku, liczba błędnych dopasowań (czerwone linie) drastycznie wzrasta, co można zaobserwować na zdjęciu 35.

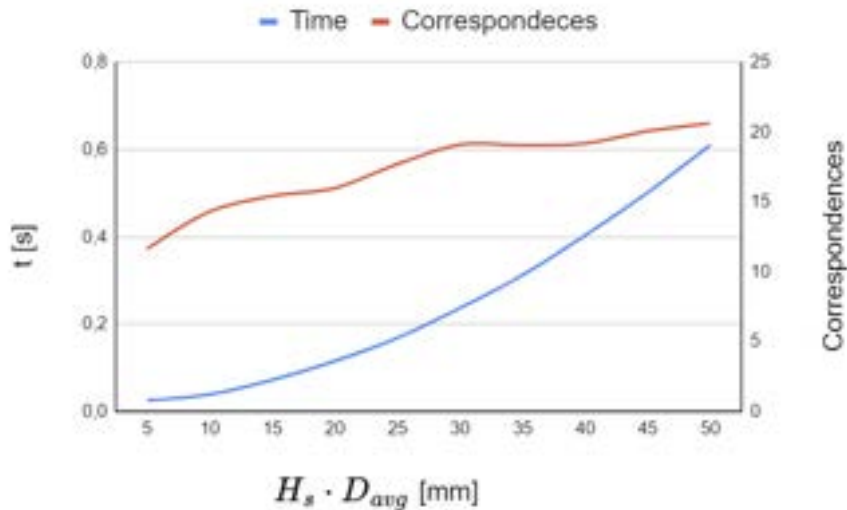
Następny parametr (T_h) jest blisko skorelowany z procesem szukania punktów charakterystycznych. Pomaga on odfiltrować punkty nie posiadające wystarczająco unikatowej cechy potrzebnej do poprawnego dopasowania. Niska wartość parametru skutkuje zdecydowanie większą liczbą błędnych dopasowań. Z drugiej jednak strony zbyt wysoka wartość T_h może wyeliminować z procesu potencjalnie poprawne korespondencje, co może całkowicie uniemożliwić złożenie dwóch chmur punktów w przypadku, gdy nie zostaną znalezione co najmniej trzy pary punktów. Wartość ustalono eksperymentalnie na $T_h = 0.2$.

Parametry promieni H_s/H_g lokalnych sfer do wyznaczania histogramów również pełnią ważną rolę w procesie ustalania korespondencji. Ich wartość znacząco wpływa na skuteczność deskryptora punktów charakterystycznych. Dodatkowo, wyższe wartości parametru oznaczają dłuższy czas wymagany do obliczenia histogramów. Zarówno zbyt mała, jak i zbyt duża

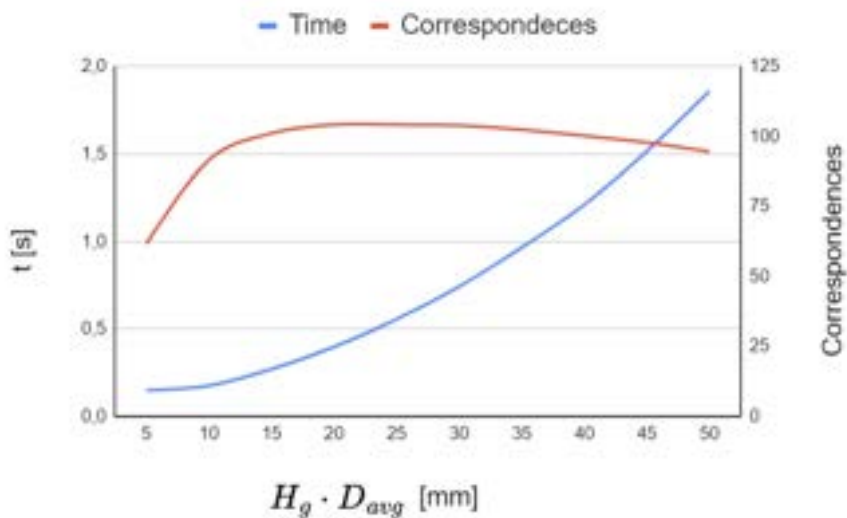


Rysunek 35: Znalezione korespondencje dla różnych przykładowych wartości parametru K_d .

wartość tego parametru negatywnie wpływa na proces szukania korespondencji, dlatego w celu ustalenia optymalnych wartości przeprowadzono metodę prób i błędów na małym podzbiore danych. Polegało to na wyznaczeniu histogramów dla różnych wartości parametru i znalezieniu korespondencji. Wartość parametru powinna być kompromisem pomiędzy liczbą znalezionych dopasowań, a czasem potrzebnym na przeprowadzenie tego procesu. Oba parametry zostały sprawdzone na podzbiore 25 par chmur punktów. Ich relacje w stosunku do czasu i znalezionych korespondencji zaprezentowano na rysunku 36 dla kształtu oraz na rysunku 37 dla gradientów intensywności.



Rysunek 36: Uśredniona relacja pomiędzy wartością promienia dla kształtu, a czasem oraz liczbą znalezionych korespondencji.



Rysunek 37: Uśredniona relacja pomiędzy wartością promienia dla gradientów intensywności, a czasem oraz liczbą znalezionych korespondencji.

Efektywność procesu szukania korespondencji wzrasta wraz ze zwiększaniem wartości promienia. Jednakże, jak zostało to już wcześniej wspomniane, czas potrzebny do znalezienia dopasowania również wzrasta. Trend ten utrzymuje się do pewnego momentu, gdzie liczba poprawnych korespondencji w przypadku parametru H_s utrzymuje się na stałym poziomie, a w przypadku H_g nawet spada. Przyjęto wartości parametrów na podstawie eksperymentu: $H_s = 30$, $30 \cdot D_{avg} \approx 20$ [mm] oraz $H_g = 15$, $15 \cdot D_{avg} \approx 10$ [mm].

Parametry N_p oraz N_T pełnią bardzo podobną rolę. Pierwszy z nich ogranicza liczbę korespondencji, a drugi obniża ilość trójek branych pod uwagę do wyznaczania transformacji 3D. Oba parametry pozwalają na zdecydowane skrócenie czasu obliczeń. W niektórych przypadkach wybór najlepszej korespondencji, bazując jedynie na podobieństwie cech, nie zawsze daje dobre wyniki dopasowania. Dlatego, aby tego uniknąć, wymagana jest analiza podzbioru kandydatów. Do osiągnięcia zadowalających rezultatów wartości tych parametrów zostały ustalone na $N_p = 500$ oraz $N_T = 500$. Wartości wszystkich parametrów zostały pogrupowane w tabeli 8.

Parametr	Wartość	Jednostka
H_b	18	-
N_{sim}	25	-
D_{avg}	0.7	[mm]
R_n	7	-
K_d	7	-
T_h	0.2	-
H_s/H_g	30/15	-
N_p	500	-
N_T	500	-

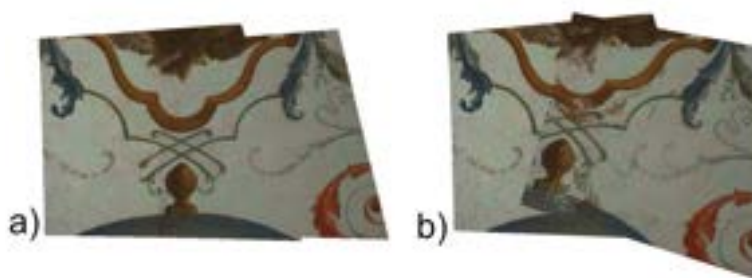
Tabela 8: Wartości parametrów metody FAMFR dla zakładanego zbioru danych.

6 Ocena skuteczności metody FAMFR

Niniejszy rozdział opisuje proces oceny zaproponowanej metody FAMFR. Przedstawione są w nim kolejno: kryterium oceny algorytmu, opis eksperymentu oraz uzyskane wyniki.

6.1 Kryterium oceny

Problem oszacowania oceny poprawnego dopasowania chmur punktów jest skomplikowany. Istniejące metryki często są niewystarczające do oceny wyniku połączenia chmur punktów, a w przypadku gęstych chmur punktów charakteryzujących się dużą liczbą zdobień i detali architektonicznych, często ocena użytkownika jest niezbędnym elementem. Przykładem może być ocena jakości poprzez minimalizację dystansu Chamfera dla płaskich fragmentów pomieszczeń odwzorowanych na chmurach punktów. Płaszczyzny mogą być wystarczająco dobrze dopasowane, jednakże tekstura na obu chmurach może być zupełnie inna (rysunek 38).



Rysunek 38: a) Poprawnie dopasowane chmury, b) Chmury posiadające podobny błąd metryki, ale błędnie połączone.

Najbezpieczniejszym, również często spotykanym sposobem jest porównywanie rozwiązania z poprawnymi, referencyjnymi wartościami. W eksperymencie do oceny wykorzystano referencyjną transformację, która powstała poprzez wstępnie połączone chmury punktów, a następnie dokładne dopasowane za pomocą algorytmu ICP.

Tak połączone pary chmur punktów wykorzystano do znalezienia punktów kontrolnych. Na początku wybrano podzbiór 1024 punktów na chmurze punktów referencyjnej. Następnie, dla każdego punktu kontrolnego zaznaczono jego odpowiednik na chmurze transformowanej, pod warunkiem, że istnieje. Punkty kontrolne wybierane są wykorzystując algorytm próbkowania do wyboru punktów charakterystycznych, aby zachować równomierny rozkład w chmurze punktów. W efekcie otrzymano zbiór korespondencji, czyli par punktów dokładnie sobie odpowiadających na wspólnym obszarze chmur punktów (rysunek 39).



Rysunek 39: Punkty kontrolne zaznaczone na części wspólnej obu chmur punktów (korespondencje).

Wyznaczone korespondencje pomiędzy punktami kontrolnymi wykorzystano do oceny finalnego dopasowania na podstawie trzech wartości: $Recall_c$, współczynnika podobieństwa cech S oraz średnio-kwadratowego błędu odległości $RMSD$.

Skuteczność algorytmu opisano jako procent poprawnych korespondencji znalezionych pomiędzy punktami kontrolnymi z pary chmur punktów ($Recall_c$). Poprawne dopasowanie dwóch punktów uznawano za prawidłowe, jeśli transformowane punkty kontrolne znajdowały się w określonej odległości od punktów w chmurze referencyjnej. Im więcej poprawnych korespondencji znajdowano, tym bardziej efektywny był algorytm.

Kolejną metryką jaką wykorzystano do oceny dopasowania był opisany wcześniej współczynnik podobieństwa S (wzór 8). Wartości te wyznaczono dla obu cech: gradientu (S_{V_g}) oraz kształtu (S_{V_s}). W celu lepszej oceny wyników wykorzystano uśrednione wartości cech w lokalnym otoczeniu. Błąd średnio-kwadratowy wyznaczono na podstawie odległości między przekształconymi punktami kontrolnymi z chmury transformowanej, a punktami utworzonymi z uśrednionych współrzędnych wszystkich punktów z chmury referencyjnej znajdujących się w lokalnym otoczeniu przekształconego punktu kontrolnego. Dodatkową metryką jaką posłużono się do analizy wyników jest dystans Chamfera [37]. Jest on często

używany jako miara poprawności dopasowania chmur punktów. Parametr ten jest obliczany przez zsumowanie kwadratów odległości pomiędzy najbliższymi, odpowiadającymi sobie punktami w dwóch chmurach punktów

6.2 Eksperyment

W przeprowadzonym eksperymencie porównano zaproponowany proces dopasowywania chmur punktów FAMFR z kilkoma klasycznymi rozwiązaniami oraz najnowocześniejszymi metodami opartymi na głębokim uczeniu. Do ewaluacji wykorzystano zbiór danych składający się ze 125 par chmur punktów pochodzących z różnych pomieszczeń i posiadających inną charakterystykę. Zbiór chmur punktów został podzielony na pięć kategorii, każda posiadająca po 25 par chmur punktów (szczegółowo zostały opisane w podrozdziale 3.3). Skuteczność metod oceniono na podstawie parametrów: $Recall_c$, współczynnika podobieństwa wektorów cech S_{V_s} oraz S_{V_g} , oraz $RMSD$ tak jak zostało opisane w podrozdziale 6.1.

Do porównania z grupą metod klasycznych wykorzystano implementację algorytmów z biblioteki PCL (Point Clouds Library [63]). Z dostępnych algorytmów zdecydowano się wykorzystać jeden do wyszukiwania punktów charakterystycznych oraz trzy deskryptory. Dodatkowo, w procesie uwzględniono klasyczny algorytm odrzucania błędnych dopasowań RANSAC [44] oraz algorytm iteracyjnego szukania precyzyjnej transformacji ICP. Cały proces można przedstawić w czterech krokach:

1. Szukanie punktów charakterystycznych: SIFT3D,
2. Wyznaczanie deskryptorów: FPFH, PFHRGB, RIFT,
3. Odrzucanie wartości odstających oraz szukanie transformacji 3D: RANSAC,
4. Wyznaczanie precyzyjnego dopasowania: ICP.

Porównanie klasycznego procesu łączenia chmur punktów oparto na trzech deskryptorach: FPFH, PFHRGB, RIFT. Wybrano algorytm FPFH ze względu na jego skuteczność, RIFT ze względu na bogatą teksturę i barwę danych wykorzystanych w testach oraz PFHRGB jako deskryptor cech oparty na kolorze i geometrii.

W przypadku metod opartych na głębokim uczeniu jednym z wyzwań oraz jednocześnie warunkiem było znalezienie działającej, udostępnionej przez autorów implementacji. Wiele z gotowych rozwiązań umieszczonych w repozytorium przystosowanych jest jedynie do

obsługi konkretnego przypadku opisanego w artykule naukowym albo tylko do odtworzenia wyników. Część repozytoriów nie była aktualizowana od publikacji, często wykorzystane w nich biblioteki są już nie wspierane, a ich wersja się znacząco zmieniła. Rzadkością jest dostępność łatwego do uruchomienia środowiska do testów na własnych danych wejściowych.

W związku z tym, w analizie skuteczności metod głębokiego uczenia wybrano pięć reprezentatywnych rozwiązań. Implementację metod wraz z modelami dla metod DCP, PointNetLK i DeepGMR zaczerpnięto z biblioteki open-source Learning3D [196]. Modele w repozytorium uczono przy użyciu zestawu danych ModelNet [197]. Z grupy nowszych metod, osiągających bardzo dobre wyniki w procesie łączenia chmur punktów zdecydowano się na: GeoTransformer [151, 198] oraz Predator [114, 115] wraz z ich oficjalnymi implementacjami. W tym przypadku modele zostały wstępnie wytrenowane przy użyciu danych 3DMath, 3DLoMath [76].

6.3 Wyniki

Wyniki porównania dopasowania dla różnych metod przedstawiono w uporządkowany sposób, zaczynając od zestawienia efektywności wszystkich algorytmów dla pięciu różnych grup chmur punktów. Pierwszy podzbiór obejmuje chmury punktów o złożonej geometrii powierzchni, reprezentujący w głównej mierze płaskorzeźby znajdujące się w skanowanych pomieszczeniach (tabela 9). Na drugi zbiór składają się płaskie chmury z bogatymi, kolorowymi zdobieniami, głównie fragmentami obrazów (tabela 10). Trzecia grupa jest zbiorem chmur punktów reprezentujących malowidła umieszczone na zakrzywionym suficie pomieszczeń (tabela 11). Czwarty podzbiór charakteryzuje się bogatymi, złotymi zdobieniami (tabela 12). Ostatnia grupa jest zbiorem 25 par chmur punktów pochodzących z pojedynczej sesji pomiarowej tego samego pomieszczenia (Gabinet al Fresco), charakteryzującego się jasną barwą oraz powtarzalnymi wzorami (tabela 13). Ostatnia tabela opisuje średnie wartości współczynników podobieństwa cech (S_{V_g} dla gradientu, S_{V_s} dla kształtu) oraz efektywności metod na wszystkich chmurach punktów wraz ze średnim czasem potrzebnym do złożenia jednej pary chmur punktów.

GRUPA I		Recall _C (%)	S_{V_g}	S_{V_s}	RMSD
Wartości referencyjne		100	0,710	0,521	0,657
Bazujące na cechach	SIFT + FPFH	37,7	0,907	0,864	4,043
	SIFT + PFHRGB	25,1	0,945	0,923	4,449
	SIFT + RIFT	14,0	0,972	0,966	4,687
Bazujące na głębokim uczeniu	DCP	3,8	0,994	0,993	4,914
	PointNetLK	7,4	0,987	0,983	4,833
	DeepGMR	14,4	0,973	0,967	4,679
	GeoTransformer	26,9	0,945	0,931	4,369
	Predator	34	0,930	0,910	4,146
	FAMFR	100	0,711	0,498	0,680

Tabela 9: Skuteczność algorytmów dla I grupy chmur punktów.

GRUPA II		Recall _C (%)	S_{V_g}	S_{V_s}	RMSD
Wartości referencyjne		100	0,420	0,672	0,195
Bazujące na cechach	SIFT + FPFH	3,3	0,993	0,990	4,924
	SIFT + PFHRGB	4,8	0,990	0,986	4,877
	SIFT + RIFT	4	0,992	0,988	4,903
Bazujące na głębokim uczeniu	DCP	14,2	0,973	0,958	4,559
	PointNetLK	13,1	0,969	0,961	4,563
	DeepGMR	68,9	0,851	0,794	2,397
	GeoTransformer	90,5	0,795	0,725	0,823
	Predator	94,2	0,781	0,714	0,867
	FAMFR	96,3	0,440	0,727	0,388

Tabela 10: Skuteczność algorytmów dla II grupy chmur punktów.

GRUPA III		Recall _C (%)	S_{V_g}	S_{V_s}	RMSD
Wartości referencyjne		100	0,508	0,457	0,141
Bazujące na cechach	SIFT + FPFH	10,8	0,972	0,965	4,713
	SIFT + PFHRGB	10,5	0,974	0,966	4,746
	SIFT + RIFT	9,7	0,975	0,969	4,751
Bazujące na głębokim uczeniu	DCP	6,4	0,985	0,979	4,842
	PointNetLK	8,8	0,977	0,972	4,771
	DeepGMR	37,6	0,897	0,870	3,893
	GeoTransformer	75,6	0,792	0,735	2,223
	Predator	84,6	0,741	0,683	1,844
	FAMFR	100	0,508	0,591	0,152

Tabela 11: Skuteczność algorytmów dla III grupy chmur punktów.

GRUPA IV		Recall _C (%)	S_{V_g}	S_{V_s}	RMSD
Wartości referencyjne		100	0,650	0,551	0,884
Bazujące na cechach	SIFT + FPFH	11,8	0,980	0,976	4,767
	SIFT + PFHRGB	10,4	0,982	0,977	4,790
	SIFT + RIFT	12	0,980	0,975	4,761
Bazujące na głębokim uczeniu	DCP	3,3	0,995	0,994	4,937
	PointNetLK	10,8	0,982	0,977	4,781
	DeepGMR	13,4	0,978	0,970	4,728
	GeoTransformer	33,3	0,941	0,916	4,293
	Predator	29,3	0,949	0,933	4,384
	FAMFR	84,4	0,734	0,644	1,849

Tabela 12: Skuteczność algorytmów dla IV grupy chmur punktów.

GRUPA V		Recall _C (%)	S_{V_g}	S_{V_s}	RMSD
Wartości referencyjne		100	0,443	0,633	0,128
Bazujące na cechach	SIFT + FPFH	0,7	0,997	0,997	4,982
	SIFT + PFHRGB	2	0,994	0,993	4,989
	SIFT + RIFT	5,7	0,975	0,979	4,765
Bazujące na głębokim uczeniu	DCP	3,1	0,993	0,990	4,927
	PointNetLK	9	0,971	0,970	4,778
	DeepGMR	5,8	0,986	0,981	4,858
	GeoTransformer	7	0,983	0,977	4,829
	Predator	8,8	0,970	0,971	4,794
FAMFR		100	0,444	0,633	0,149

Tabela 13: Skuteczność algorytmów dla V grupy chmur punktów.

ŚREDNIA		Recall _C (%)	S_{V_g}	S_{V_s}	RMSD	Time
Wartości referencyjne		100	0,546	0,567	0,401	≈ 200 [s]
Bazujące na cechach	SIFT + FPFH	8,2	0,983	0,978	4,823	158 [s]
	SIFT + PFHRGB	10	0,978	0,971	4,783	132 [s]
	SIFT + RIFT	6,4	0,985	0,982	4,837	47 [s]
Bazujące na głębokim uczeniu	DCP	6,2	0,988	0,983	4,836	239 [ms]
	PointNetLK	9,8	0,977	0,973	4,745	295 [ms]
	DeepGMR	28	0,937	0,916	4,111	137 [ms]
	GeoTransformer	46,7	0,891	0,857	3,307	345 [ms]
	Predator	50,2	0,874	0,842	3,207	1868 [ms]
FAMFR		96,1	0,567	0,619	0,644	17 [s]

Tabela 14: Uśredniona skuteczność algorytmów dla wszystkich chmur punktów.

7 Dyskusja

Dyskusję podzielono na dwie części. Pierwsza dotyczy analizy wyzwań i problemów jakie napotkano podczas przeprowadzania procesu łączenia chmur punktów metodą FAMFR. W drugiej części omówiono wyniki eksperymentu oraz przedyskutowano różnice w wynikach względem innych metod.

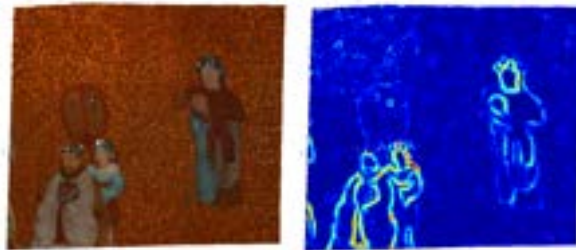
7.1 Wyzwania oraz ograniczenia metody FAMFR

Podczas eksperymentu napotkano kilka często powtarzających się błędów w procesie łączenia chmur punktów. Jednym ze źródeł problemów jest mała powierzchnia części wspólnej obu chmur punktów występująca w niektórych analizowanych przypadkach. Przykładem może być para chmur punktów (Rysunek 40), gdzie chmury pokrywają się w bardzo małym stopniu. Również bardzo mała złożoność geometryczna kształtu oraz mało różnorodny kolor części wspólnej może w znaczący sposób utrudnić proces łączenia chmur punktów. Warto zauważyć, że zaproponowana metoda FAMFR efektywnie zadziałała nawet w tak wymagającym i trudnym przypadku.



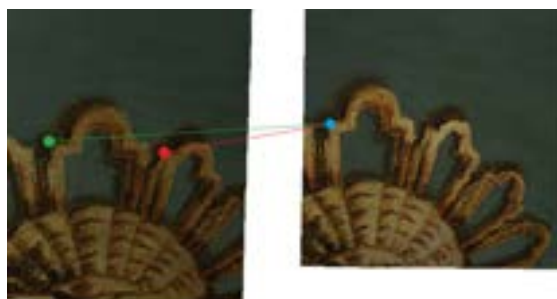
Rysunek 40: Para chmur punktów z małym fragmentem części wspólnej.

Innym wymagającym scenariuszem, jest łączenie chmur punktów posiadających szczegółową teksturę oraz częściowo odblaskową powierzchnię. Tworzy ona wiele charakterystycznych fragmentów chmury, które zdecydowanie zwiększają możliwość wystąpienia błędnego dopasowania. W przypadku źle dobranych parametrów otrzymuje się albo zbyt dużo błędnych korespondencji, albo wspólne fragmenty obu chmur nie posiadają wystarczającej liczby poprawnych punktów charakterystycznych. W testowanym zbiorze par chmur punktów znajdowały się cztery takie przypadki (dla przykładu, rysunek 41). Z tego podzbioru, zaproponowana metoda poprawnie połączyła trzy z czterech par chmur punktów.



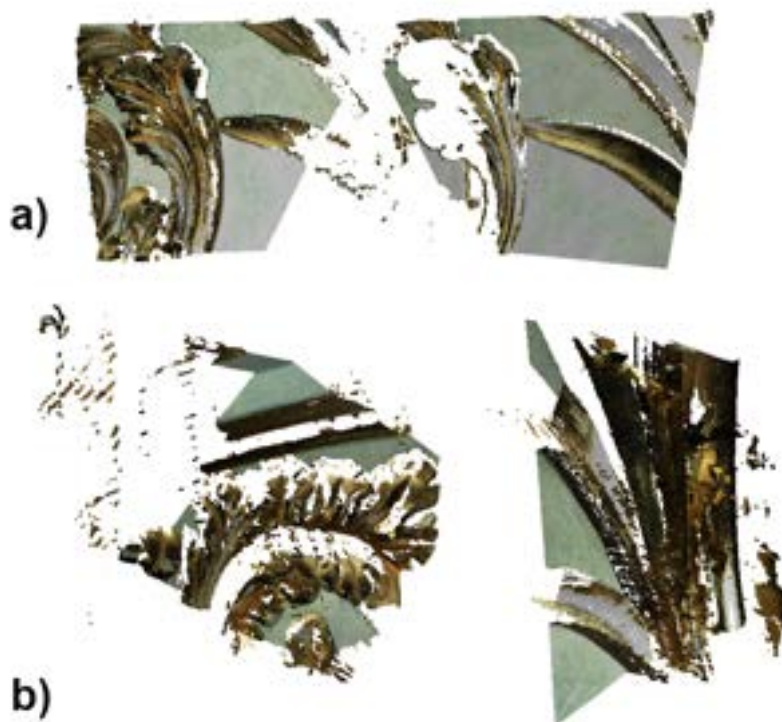
Rysunek 41: Chmury punktów z bogato zdobioną, częściowo odblaskową teksturą. Po prawej wyznaczone wartości gradientów.

Błędne dopasowania wykrywano również podczas łączenia chmur punktów posiadających powtarzalne wzory w postaci np. zdobień, kolorów lub tekstury. Wyznaczone deskryptory dla punktów na obu chmurach opisujące fragmenty z powtarzalną cechą posiadały wiele teoretycznie poprawnych odpowiedników. Często bazując na współczynniku podobieństwa najlepsza korespondencja dla danego punktu nie była tą właściwą. W takich przypadkach to błędne korespondencje były głównym powodem źle wyznaczonej transformacji. Przykład takiej chmury wraz z błędnymi korespondencjami zaprezentowano na rysunku 42.



Rysunek 42: Przykład błędnie wyznaczonych korespondencji: czerwoną linią zaznaczono błędne dopasowanie, zieloną poprawne.

Czwarta grupa chmur punktów była najbardziej wymagająca w procesie łączenia (tabela 12). Obecność wielu refleksyjnych, połączonych zdobień znacząco utrudniała dopasowywanie chmur punktów poprzez wysoki poziom szumu pomiarowego. Światło odbijające się od złotych powierzchni powodowało wiele problemów w rekonstrukcji danego fragmentu, takich jak: słabe odwzorowanie geometrii powierzchni oraz koloru, gromadzenie się dużej liczby punktów szumowych poza powierzchnią obiektu lub tworzenie nie istniejących fragmentów powierzchni obiektu. Przykłady tych problemów zaprezentowano na rysunku 43.



Rysunek 43: Szum pomiarowy napotkany przy łączeniu chmur punktów z podgrupy IV: a) brakująca rekonstrukcja powierzchni złotego liścia z prawej chmury punktów, b) dwie chmury punktów z wysokim poziomem szumu pomiarowego.

7.2 Analiza wyników

Jak zostało to zaprezentowane w tabeli 14, FAMFR przewyższa wszystkie inne testowane algorytmy pod kątem efektywności na tym konkretnym zbiorze danych.

Pierwsza grupa metod klasycznych, opartych na cechach, wykazywała bardzo słabe wyniki. Chociaż metody łączenia chmur punktów oparte na cechach charakterystycznych mają tę zaletę, że wymagają jedynie modelu 3D obiektu, procesy obliczeń wartości wektorów cech oraz znajdowania korespondencji są wymagające obliczeniowo. Wszystkie metody oparte na cechach zostały poddane ocenie efektywności za pomocą kilku zbiorów parametrów zdefiniowanych przez użytkownika w celu dokładnego porównania. Algorytm SIFT3D został skonfigurowany do wykrywania trzech różnych ilościowo zestawów punktów kluczowych: 800, 3500 i 6000. W podobny sposób poddano analizie parametry deskryptora w celu znalezienia zestawu, który dawał najlepsze rezultaty. Uwzględniono różne wartości parametru odpowiadającego za promień lokalnego otoczenia wymaganego do obliczenia cech, dokładnie: 5, 10 i 20 milimetrów dla metod FPFH i RIFT. Ze względu na długi czas wymagany (około 30 minut na pojedynczą chmurę) do obliczenia cechy PFHRGB, dla tego deskryptora zastosowano stałą wartość promienia równą 5 milimetrów. Takie podejście umożliwiło bardziej rygorystyczną i solidną ocenę metod oraz pomogło zidentyfikować optymalne ustawienia parametrów dla rzetelnego porównania. W tabelach uwzględniono i zaprezentowano średnie czasy oszacowania, błędy współczynnika podobieństwa oraz skuteczność dopasowywania jedynie dla ustawień parametrów, które dawały najlepsze wyniki.

Zgodnie z założeniami, jakimi kierowano się przy wyborze metod, najlepszą wydajność osiągnięto przy użyciu metody PFHRGB. Uwzględnia ona komplet informacji z obu cech opisujących chmurę: geometrię oraz kolor. Algorytm FPFH, który bierze pod uwagę tylko informacje o kształcie, uzyskał niższą skuteczność. Wynika to z niezdolności algorytmu do wyodrębnienia wystarczająco charakterystycznych cech do identyfikacji poprawnych korespondencji w przypadku obiektów o płaskich powierzchniach lub powierzchniach o stałej krzywiznie. Algorytm RIFT, który analizuje wektory gradientów w lokalnym otoczeniu osiągnął również słabe wyniki. Jest to prawdopodobnie spowodowane błędami tekstury i kolorów spowodowanymi rozbłyskami i odbiciami lustrzanymi występującymi na etapie skanowania powierzchni pomieszczeń. Ponadto wszystkie wymienione deskryptory cech obciążone są również błędami wynikającymi z procesu szukania punktów kluczowych za pomocą algorytmu SIFT3D. Zbiór poprawnych korespondencji opartych na tak wyznaczonych

punktach może być niewielki lub zerowy. Nie ma pewności, że punkty znalezione na chmurze transformowanej mają swoje odpowiedniki na chmurze stabilnej. Wszystko to może prowadzić do błędnego oszacowania finalnej transformacji. Należy zauważyć, że wszystkie metody oparte na cechach mają pewne wady. Są matematycznie skomplikowane, ciężkie obliczeniowo i wrażliwe na dostrajanie parametrów. Często wymogiem jest posiadanie specjalistycznej wiedzy do określenia optymalnych wartości parametrów dla zadanego zestawu danych.

Jedna z pierwszych metod opartych na głębokim uczeniu (DCP) osiągnęła bardzo niską skuteczność. Głównym powodem są dane wykorzystane w rozprawie doktorskiej. Algorytm opiera się na dokładnym odwzorowaniu chmur referencyjnej i transformowanej. Założono, że każdy punkt w jednej chmurze posiada dokładnie jeden swój odpowiednik w drugiej. Założenie to jest nie zawsze odpowiednie w przypadku scenariuszy opartych na rzeczywistych obiektach. Również szum pomiarowy w danych negatywnie wpływał na działanie algorytmu, ze względu na złożoną konstrukcję modelu oraz fakt, że wykorzystane lokalne cechy opisujące geometrię są szczególnie wrażliwe na ten rodzaj błędu. PointNetLK, który był uczony do wykrywania cech charakterystycznych dla określonych kategorii obiektów, nie rozpoznał wartościowych cech w chmurach punktów użytych w tym badaniu. Prowadziło to do znajdowania dużej liczby błędnych punktów wiążących pomiędzy parą chmur punktów. Dodatkowo, podczas procesu dopasowywania, algorytm z łatwością może wpaść w lokalne minima. Metoda DeepGMR oszacowała zgodność między wszystkimi punktami i wszystkimi składnikami w utajonym modelu mieszaniny Gaussa (GMM), dzięki czemu wynik rejestracji jest niezmienny w stosunku do gęstości geometrii wejściowych chmur punktów. Metoda zakłada idealne dopasowanie między dwoma rozkładami chmur punktów, co nie jest prawdą w przypadku chmur punktów użytych w tym badaniu. Zestaw danych posiada pewien poziom szumu pomiarowego oraz inne niepewności związane z rekonstrukcją geometrii i odwzorowaniem koloru. Konsekwencją tego była słaba efektywność dopasowywania chmur punktów. Mimo wszystko, algorytm osiągnął lepsze wyniki niż DCP oraz PointNetLK.

Dwie nowsze metody, GeoTransformer i Predator wymagały przeskalowania danych z powodu wymagań dotyczących znacznego wykorzystania pamięci oraz dużych kosztów obliczeniowych. GeoTransformer, oparty na geometrycznych cechach niezależnych od położenia, osiągnął znacznie lepsze wyniki niż wcześniej opisane metody głębokiego uczenia. Metoda ta wykorzystuje wyuczone cechy geometryczne wraz z analizą zgodności przestrzennej, co czyni ją bardziej niezawodną w przypadkach chmur punktów, które

nakładają się w bardzo niskim stopniu. Głównym źródłem problemów metody GeoTransformer są hiperparametry wykorzystywane do kontrolowania czułości zmian odległości pomiędzy punktami i kątów trójkątów. Muszą być one precyzyjnie wybrane dla różnych zestawów danych. Predator z kolei jest architekturą sieci neuronowej uczącą się wykrywać obszar wspólny dwóch nie połączonych chmur punktów i wykorzystywać ten region do próbkowania superpunktów. Mimo, że metoda Predator ma ograniczenia w przypadkach, w których gęstość punktów jest bardzo nierówna, wykazano, że jego zdolność do ustalania punktów istotnych dla procesu dopasowywania znacząco zwiększa jego wydajność. Predator osiągnął najlepsze wyniki w eksperymencie w porównaniu z innymi metodami głębokiego uczenia.

Biorąc pod uwagę złożoność obliczeniową i czas wykonywania algorytmu, zaproponowana metoda wykazuje lepsze wyniki w porównaniu z testowanymi klasycznymi metodami opartymi na cechach. Z drugiej jednak strony ustępuje pod względem szybkości metodom głębokiego uczenia. Dzieje się tak jednak w przypadku, gdy nie bierzemy pod uwagę czasu potrzebnego na uczenie sieci neuronowej. Dodatkowo nie do pominięcia jest czas przygotowania zbioru uczącego wraz z często wymaganym wstępnym przetwarzaniem chmur punktów. Czas trwania procesu uczenia sieci w typowej metodzie głębokiego uczenia może się znacznie różnić w zależności od wielu czynników, takich jak złożoność sieci, ilość danych wykorzystywanych do uczenia, rodzaju wykonywanego zadania oraz dostępnych zasobów obliczeniowych. Trening modelu głębokiego uczenia jest procesem iteracyjnym i może wymagać wielu etapów szkolenia, oceny i dostrajania hiperparametrów, aby osiągnąć optymalne wyniki. W niektórych przypadkach szkolenie może zająć tylko kilka minut, podczas gdy w innych może zająć kilka dni lub tygodni. Najnowocześniejsze modele przetwarzania obrazu komputerowego oraz modele języka naturalnego (ang. Natural Language Processing, NLP [199]) mogą wymagać kilkudniowego, a nawet tygodniowego procesu uczenia wykorzystując wysokiej jakości sprzęt do obliczeń. Dlatego podanie konkretnego czasu uczenia, który ma zastosowanie do wszystkich scenariuszy, jest trudne. W przypadku metod głębokiego uczenia, nacisk kładzie się często na metryki wydajności modelu, takie jak dokładność lub wskaźnik F1 [200], a nie na czas uczenia. W założeniu nie jest on istotnym czynnikiem oceny jakości modelu, ponieważ jest to koszt jednorazowy i nie wpływa na wydajność zaproponowanej metody.

Jak zostało to już wiele razy podkreślone, chmury punktów mogą się znacząco różnić

w zależności od typów obiektów dziedzictwa kulturowego. W wielu przypadkach dane z konkretnego obiektu są unikatowe i wyuczony model może nie działać skutecznie w przypadku innego typu obiektu. Często, konieczne jest utworzenie zupełnie nowego zestawu danych szkoleniowych, gdyż sam proces douczania może być niewystarczający. Dla konkretnych pomieszczeń DK wykorzystanych w niniejszej rozprawie, przygotowanie zbioru uczącego z dbałością o szczegóły wynikające z bogatej i zróżnicowanej aranżacji wnętrza jest znaczącym wyzwaniem. Zdolność metod głębokiego uczenia do efektywnego działania na nowych danych, które nie zostały uwzględnione podczas procesu uczenia, jest często ograniczona. Dzieje się tak, ponieważ wyuczony model może nie być dobrze uogólniony na nowe i zróżnicowane obiekty.

Istniejące implementacje metod opartych na głębokim uczeniu często nie zapewniają łatwego w obsłudze programu do szybkiego uruchomienia procesu na własnych danych wejściowych. Zazwyczaj są one przeznaczone dla badaczy i zaawansowanych użytkowników posiadających wiedzę dotyczącą sieci neuronowych i złożoności metod głębokiego uczenia. Metody te często obejmują złożone architektury sieci neuronowych i wymagają dużych ilości danych do uczenia. Ponadto metody te są stosunkowo nowe w porównaniu z metodami głębokiego uczenia opartymi na obrazach 2D, a badania w tej dziedzinie wciąż trwają. W rezultacie wdrożenie tych metod jest trudniejsze w stosunku do tradycyjnych metod, które zostały już wnikliwie przebadane i istnieje wiele łatwo dostępnych implementacji. Metody głębokiego uczenia często wymagają znacznych zasobów obliczeniowych, co dodatkowo może zwiększać złożoność korzystania z tych metod. Pomimo tych wyzwań, nadal podejmowane są wysiłki w celu opracowania bardziej przyjaznych dla użytkownika implementacji metod łączenia chmur punktów opartych na głębokim uczeniu. Należy jednak pamiętać, że te metody nie zawsze są najlepszym rozwiązaniem problemów związanych z dopasowywaniem chmur punktów. W wielu przypadkach, tradycyjne metody nadal są bardziej odpowiednie i stanowią nieodzowny element procesów łączenia chmur punktów.

8 Podsumowanie

W niniejszej rozprawie doktorskiej wykorzystano chmury punktów reprezentujące wnętrza pomieszczeń Muzeum Pałacu Króla Jana III w Wilanowie charakteryzujące się złożonymi i unikalnymi cechami powierzchni. Istniejące metody nie rozwiązują problemu dopasowywania chmur punktów na tych danych, dlatego istnieje potrzeba implementacji efektywnego i wydajnego rozwiązania. Zaproponowana metoda FAMFR poprawnie łączy chmury punktów poprzez właściwe wykorzystanie dwóch różnych cech: V_g , która zawiera informacje o gradiencie intensywności oraz V_s , która opisuje geometryczną zależność między sąsiednimi punktami a ich wektorami normalnymi. Wykorzystanie obu tych cech do wyznaczenia punktów charakterystycznych oraz ich deskryptorów za pomocą histogramów w skuteczny sposób umożliwia wstępne dopasowanie chmur punktów. Angażując w końcowym etapie algorytm ICP otrzymuje się kompletne narzędzie, umożliwiające dokładną estymację transformacji 3D pomiędzy dwoma chmurami punktów. Wyniki eksperymentu potwierdzają skuteczność metody FAMFR na wszystkich grupach chmur punktów, osiągając poprawę efektywności łączenia chmur punktów o około 80% w porównaniu z tradycyjnymi metodami opartymi na cechach i około 45% w porównaniu z metodami opartymi na głębokim uczeniu.

Algorytm FAMFR, mimo udokumentowanej skuteczności w łączeniu chmur punktów posiada pewne ograniczenia i słabości. Eksperyment obejmował zestaw wymagających danych posiadających złożone powierzchnie o wysokim połysku (chmury punktów z grupy IV, 43), gdzie napotkano największe wyzwania. Mimo, że zaproponowana metoda poprawnie zadziałała w większości przypadków, to w porównaniu z innymi scenariuszami osiągnęła stosunkowo niższą skuteczność. Inne ograniczenie jakie zaobserwowano podczas eksperymentu zachodziło w przypadku pary skanów o niewielkim wspólnym regionie (rysunek 40). Znalezienie prawidłowych korespondencji w takim przypadku jest trudne, zwłaszcza gdy taki region dodatkowo posiada ograniczone informacje o kolorze, teksturze lub geometrii. Punkty wykorzystywane w zaproponowanej metodzie do wstępnego łączenia chmur punktów są równomiernie próbkowane na powierzchni chmury, dlatego im mniejszy jest region w którym chmury się nakładają, tym mniejszy jest zbiór potencjalnych korespondencji. Ponadto, taki zbiór jest poddany procesowi filtracji, a częściowo odbłaskowe, przypominające szum pomiarowy tekstury (rysunek 41) oraz powtarzające się elementy zdobień (rysunek 42) tworzą wiele błędnych dopasowań. Gdy nie zostaną prawidłowo wykryte i odrzucone, mogą

zostać błędnie zakwalifikowane jako poprawne korespondencje, co może doprowadzić do nieprawidłowo wyznaczonej wstępnej transformacji 3D.

8.1 Realizacja celów pracy

Celem rozprawy doktorskiej było opracowanie skutecznej metody automatycznej rekonstrukcji obiektów dziedzictwa kulturowego. Zaproponowano autorski sposób łączenia chmur punktów, będących efektem skanowania pomieszczeń obiektów dziedzictwa kulturowego wraz z implementacją algorytmów oraz ich weryfikacją. Zadanie to zostało podzielone na cele naukowe oraz cel aplikacyjny.

8.1.1 Realizacja celów naukowych

Zrealizowany cel obejmował badania i opracowanie metody dopasowującej chmury punktów bazującej na cechach geometrycznych oraz intensywności wyznaczonej z koloru na podstawie lokalnego otoczenia w danym punkcie oraz jej implementację. Metoda spełnia warunek dotyczący uzależnienia parametrów wejściowych od średniej odległości między punktami, co zostało dokładniej opisane w podrozdziale 5.6. Rezultat działania musiał spełniać określone warunki na podstawie oszacowanych metryk badających podobieństwo par chmur punktów:

- kontrolny błąd dopasowania $RMSD < 1.5 \cdot D_{avg} \approx 1[mm]$,
- różnica wyliczonego dystansu Chamfera [37] pomiędzy wynikiem, a wartością referencyjną $D_{CHd} < 3 \cdot D_{avg} \approx 2.1[mm]$,
- różnica współczynnika podobieństwa $S(8)$ pomiędzy wynikiem, a wartością referencyjną $S_{Gd} < 0.1$ & $S_{Sd} < 0.15$ (15% maksymalnej wartości dla cechy kształtu Sd i gradientów Gd).

Dodatkowo, algorytm musiał być zdolny do poprawnego połączenia danych wejściowych o różnej charakterystyce powierzchni pochodzących z pomieszczeń obiektów dziedzictwa kulturowego. Są to między innymi chmury punktów reprezentujące płaskie ściany z bogatymi zdobieniami i malowidłami, płaskorzeźby oraz obiekty o złożonym kształcie, przy uwzględnieniu obecności licznych złożonych i błyszczących dekoracji. Ocena rezultatu działania metody została zaprezentowana w tabeli 15, gdzie znajdują się wyniki wymienionych metryk dla każdego zbioru danych oraz ich uśrednione wartości. Określone

warunki zostały spełnione, a w niektórych przypadkach zaproponowany algorytm osiągnął lepsze wyniki niż manualne dopasowanie (wartości ujemne w tabeli).

FAMFR	$RMSD$	D_{CHd}	S_{Gd}	S_{Sd}
GRUPA I	0,0231	0,06	0,0002	-0,023
GRUPA II	0,193	0,11	0,020	0,054
GRUPA III	0,0117	2,02	0,0004	0,134
GRUPA IV	0,965	0,54	0,084	0,093
GRUPA V	0,022	0,32	0,002	-0,0002
ŚREDNIA	0,2429 < 1,0	0,61 < 2,1	0,0212 < 0,15	0,052 < 0,15

Tabela 15: Zestawienie metryk podobieństwa wyników dla każdej z grup metodą FAMFR.

8.1.2 Realizacja celu aplikacyjnego

Cel aplikacyjny został spełniony poprzez skuteczne wykorzystanie zaproponowanej metody do efektywnego łączenia dużej liczby chmur punktów będących efektem kampanii skanowania 3D pomieszczeń wewnątrz Muzeum Pałacu Króla Jana III w Wilanowie (więcej szczegółów w podrozdziale 1.2). Metoda FAMFR okazała się skutecznym narzędziem praktycznym do dopasowywania chmur punktów o wysokiej rozdzielczości dla zadanego zbioru danych z pomieszczeń obiektów dziedzictwa kulturowego. Pozwala na szybkie i efektywne łączenie chmur punktów, znacznie ułatwiając rekonstrukcję dużych modeli 3D obiektów dziedzictwa kulturowego. Czas potrzebny na manualną orientację 50 chmur punktów wynosi około 10 000 sekund, podczas gdy zaproponowane rozwiązanie skraca ten czas o ponad 90%, do zaledwie 850 sekund. Dodatkowo metoda spełnia warunek jakim jest automatyzacja oraz parametryzacja poprzez zaproponowanie niewielkiej liczby sterujących parametrów wejściowych, które są uzależnione od danych wejściowych. Metoda jest łatwa w użyciu i gotowa do zaimplementowania w innych procesach wymagających precyzyjnego dopasowywania chmur punktów. Efekty prac wykorzystujące niniejszą metodę dla konkretnych pomieszczeń są zaprezentowane na rysunkach: Antygabinet Królowej (rysunek 44), Gabinet Chiński Króla (rysunek 45), Garderoba Króla (rysunek 46), Gabinet al Fresco (rysunek 47).



Rysunek 44: Pełny model południowej i wschodniej ściany Antygabinetu Królowej.



Rysunek 45: Zbliżenie na południowo-zachodnie naroże Gabinetu Chińskiego Króla.



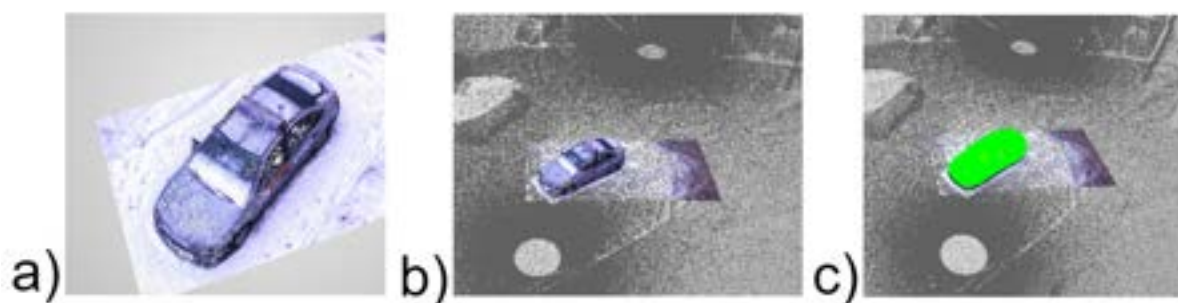
Rysunek 46: Wschodnia ściana Garderoby Króla.



Rysunek 47: Północna ściana Gabinetu al Fresco.

8.2 Potencjalne dodatkowe zastosowania

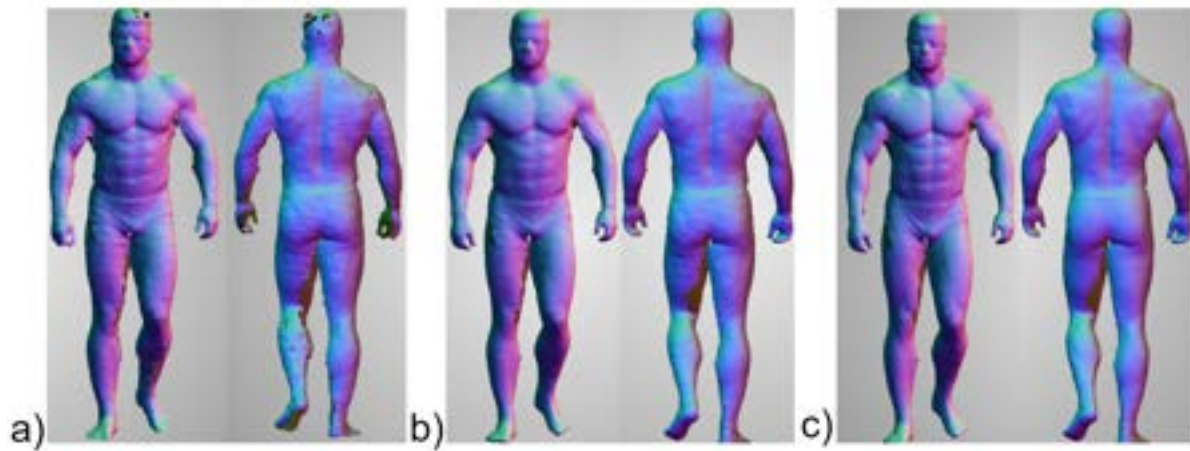
Metoda FAMFR znalazła swoje zastosowanie również w projekcie analizującym infrastrukturę krytyczną. Projekt "System automatycznej detekcji zagrożeń infrastruktury krytycznej oparty na multimodalnym układzie pomiarowym" powstał w ramach grantu NCBiR oraz przy udziale firmy GISS [201]. Jednym z elementów projektu był moduł analizy porównawczej danych 3D zarejestrowanych po określonym czasie z danymi referencyjnymi obiektu infrastruktury krytycznej. Jej celem była identyfikacja potencjalnych zagrożeń wynikających np. z wykrytego obiektu w postaci samochodu bądź budynku w strefie ochronnej. Analiza porównawcza wymagała połączenia chmur punktów, referencyjnych danych oraz najnowszych pomiarów. Rezultat oraz wyniki działania przedstawiono na rysunku 48.



Rysunek 48: Proces detekcji różnicowej w projekcie GISS: a) wykryty obiekt, b) połączenie z modelem referencyjnym (szara chmur punktów), c) znaleziony nowy obiekt (zielony kolor).

Metody zaimplementowane w ramach FAMFR znalazły swoje zastosowanie w start-upie Mnemosis [202]. Jednym z produktów jest tworzenie cyfrowych bliźniaków wykorzystując multimodalne dane pochodzące z trzech metod pomiarowych: SL (ang. Structured Light), SfS (ang. Shape from Silhouette) oraz SfM (ang. Structure from Motion). Wynikiem pomiaru wymienionymi metodami były chmury punktów, które należało połączyć w celu stworzenia dokładnego modelu skanowanego człowieka. Każda z metod posiadała pożądane cechy, które w odpowiedni sposób ulepszały końcowy model 3D. Bazowy model, w postaci chmury punktów SfM, był uzupełniany modelem SfS do uzupełnienia powierzchni geometrii i usprawnienia rekonstrukcji poszczególnych fragmentów, a następnie do całości dodawany był model SL do uzyskania zdecydowanie lepszej jakości odwzorowania powierzchni.

Połączony model człowieka podzielony na poszczególne etapy zaprezentowano na rysunku 49.



Rysunek 49: Etapy wzbogacania modelu człowieka w projekcie Mnemosis: a) bazowy model SfM, b) połączenie z modelem SfS, uzupełnione fragmenty modelu (np. głowa) oraz lepiej odwzorowane dłonie, c) połączenie z modelem SL, wygładzenie powierzchni i uwypuklenie szczegółów.

8.3 Przyszłe prace

Przyszłe prace będą przede wszystkim obejmować ocenę działania zaproponowanej metody na innych typach danych. Może to pomóc w ocenie skalowalności oraz efektywności algorytmu FAMFR. Dodatkowo, można rozważyć dalszą optymalizację kodu w celu skrócenia czasu obliczeń. Może to obejmować badania zarówno związane ze zrównoleglaniem procesu obliczeń, wykorzystując jak największą liczbę wątków procesora CPU jak i optymalizację wykorzystania pamięci oraz odpowiednich struktur danych.

Inną potencjalną drogą rozwoju jest wykorzystanie zaproponowanych cech do stworzenia modelu opartego na głębokim uczeniu. Model może posłużyć zarówno do opisu punktów charakterystycznych jak i do filtracji i odrzucania błędnych punktów wiążących. Może to potencjalnie poprawić dokładność i niezawodność procesu dopasowywania. W tym przypadku należałoby skupić się na potencjalnej możliwości transferu uczenia (ang. transfer learning) na nowe zestawy danych oraz odpowiednio przygotować zbiór danych uczących. Następnym krokiem powinna być ewaluacja implementacji głębokiego uczenia korzystając z dostępnych referencyjnych zestawów chmur punktów służących powszechnie do oceny jakości dopasowań (np. KITTY, 3DMatch, 3DLowMatch). Pozwoli to na dokładniejszą ewaluację metody w porównaniu z istniejącymi rozwiązaniami.

9 Bibliografia

Literatura

- [1] Z. Dong, F. Liang, B. Yang, Y. Xu, Y. Zang, J. Li, Y. Wang, W. Dai, H. Fan, J. Hyypä *et al.*, “Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 163, pp. 327–342, 2020.
- [2] N. Xu, R. Qin, and S. Song, “Point cloud registration for lidar and photogrammetric data: A critical synthesis and performance analysis on classic and deep learning algorithms,” *ISPRS Open Journal of Photogrammetry and Remote Sensing*, p. 100032, 2023.
- [3] S. Tuttas, A. Braun, A. Borrmann, and U. Stilla, “Acquisition and consecutive registration of photogrammetric point clouds for construction progress monitoring using a 4d bim,” *PFG–journal of photogrammetry, remote sensing and geoinformation science*, vol. 85, no. 1, pp. 3–15, 2017.
- [4] J. Shao, W. Zhang, N. Mellado, P. Grussenmeyer, R. Li, Y. Chen, P. Wan, X. Zhang, and S. Cai, “Automated markerless registration of point clouds from tls and structured light scanner for heritage documentation,” *Journal of Cultural Heritage*, vol. 35, pp. 16–24, 2019.
- [5] Q. Cheng, P. Sun, C. Yang, Y. Yang, and P. X. Liu, “A morphing-based 3d point cloud reconstruction framework for medical image processing,” *Computer methods and programs in biomedicine*, vol. 193, p. 105495, 2020.
- [6] Z. Sharifisoraki, A. Dey, R. Selzler, M. Amini, J. Green, S. Rajan, and F. Kwamena, “Monitoring critical infrastructure using 3d lidar point clouds,” *IEEE Access*, vol. 11, pp. 314–336, 2022.
- [7] L.-P. Bergé, N. Aouf, T. Duval, and G. Coppin, “Generation and vr visualization of 3d point clouds for drone target validation assisted by an operator,” in *2016 8th Computer Science and Electronic Engineering (CEECE)*. IEEE, 2016, pp. 66–70.
- [8] M. Canciani, C. Falcolini, M. Saccone, and G. Spadafora, “From point clouds to architectural models: algorithms for shape reconstruction,” *The International Archives*

- of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, pp. 27–34, 2013.
- [9] D. Bolkas, J. Chiampi, J. Chapman, and V. F. Pavill, “Creating a virtual reality environment with a fusion of suas and tls point-clouds,” *International journal of image and data fusion*, vol. 11, no. 2, pp. 136–161, 2020.
- [10] E. Grilli, E. Farella, A. Torresani, and F. Remondino, “Geometric features analysis for the classification of cultural heritage point clouds,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 541–548, 2019.
- [11] T. Bakirman, B. Bayram, B. Akpınar, M. F. Karabulut, O. C. Bayrak, A. Yigitoglu, and D. Z. Seker, “Implementation of ultra-light uav systems for cultural heritage documentation,” *Journal of Cultural Heritage*, vol. 44, pp. 174–184, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1296207419303747>
- [12] R. Zlot, M. Bosse, K. Greenop, Z. Jarzab, E. Juckes, and J. Roberts, “Efficiently capturing large, complex cultural heritage sites with a handheld mobile 3d laser mapping system,” *Journal of Cultural Heritage*, 12 2013.
- [13] R. Boboc, F. Gîrbacia, C. Postelnicu, and T. Gîrbacia, “Evaluation of using mobile devices for 3d reconstruction of cultural heritage artifacts,” in *VR Technologies in Cultural Heritage*. Springer International Publishing, 2019, pp. 46–59.
- [14] K. Kingsland, “Comparative analysis of digital photogrammetry software for cultural heritage,” *Digital Applications in Archaeology and Cultural Heritage*, vol. 18, p. e00157, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212054820300564>
- [15] A. Bocheńska, J. Markiewicz, and S. Łapiński, “The combination of the image and range-based 3d acquisition in archaeological and architectural research in the royal castle in warsaw,” *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 177–184, 2019.

- [16] S. Saha, J. Martusewicz, N. L. Streeton, and R. Sitnik, “Segmentation of change in surface geometry analysis for cultural heritage applications,” *Sensors*, vol. 21, no. 14, p. 4899, 2021.
- [17] S. Saha, A. Siatou, A. Mansouri, and R. Sitnik, “Supervised segmentation of rti appearance attributes for change detection on cultural heritage surfaces,” *Heritage Science*, vol. 10, no. 1, pp. 1–15, 2022.
- [18] E. Stylianidis, “Cipa - heritage documentation: 50 years: Looking backwards,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W14, pp. 1–130, 2019. [Online]. Available: <https://isprs-archives.copernicus.org/articles/XLII-2-W14/1/2019/>
- [19] P. Tysiac, A. Sieńska, M. Tarnowska, P. Kedziorowski, and M. Jagoda, “Combination of terrestrial laser scanning and uav photogrammetry for 3d modelling and degradation assessment of heritage building based on a lighting analysis: case study—st. adalbert church in gdansk, poland,” *Heritage Science*, vol. 11, no. 1, p. 53, Mar 2023. [Online]. Available: <https://doi.org/10.1186/s40494-023-00897-5>
- [20] M. Guo, M. Sun, D. Pan, G. Wang, Y. Zhou, B. Yan, and Z. Fu, “High-precision deformation analysis of yingxian wooden pagoda based on uav image and terrestrial lidar point cloud,” *Heritage Science*, vol. 11, no. 1, p. 1, Jan 2023. [Online]. Available: <https://doi.org/10.1186/s40494-022-00833-z>
- [21] G. R. Bent, D. Pfaff, M. Brooks, R. Radpour, and J. Delaney, “A practical workflow for the 3d reconstruction of complex historic sites and their decorative interiors: Florence as it was and the church of orsanmichele,” *Heritage Science*, vol. 10, no. 1, p. 118, Jul 2022. [Online]. Available: <https://doi.org/10.1186/s40494-022-00750-1>
- [22] A. Tobiasz, J. Markiewicz, S. Łapiński, J. Nickel, P. Kot, and M. Muradov, “Review of methods for documentation, management, and sustainability of cultural heritage. case study: Museum of king jan iii’s palace at wilanów,” *Sustainability*, vol. 11, no. 24, p. 7046, 2019.
- [23] E. Abbate, G. Sammartano, and A. Spanò, “Prospective upon multi-source urban scale data for 3d documentation and monitoring of urban legacies,” *The International Archives*

- of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 11–19, 2019.
- [24] S. Del Pozo, P. Rodríguez-González, L. Sánchez-Aparicio, A. Muñoz-Nieto, D. Hernández-López, B. Felipe-García, and D. González-Aguilera, “Multispectral imaging in cultural heritage conservation,” *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 155–162, 2017.
- [25] S. Yang, S. Xu, and W. Huang, “3d point cloud for cultural heritage: A scientometric survey,” *Remote Sensing*, vol. 14, no. 21, 2022. [Online]. Available: <https://www.mdpi.com/2072-4292/14/21/5542>
- [26] E. Grilli, E. Farella, A. Torresani, and F. Remondino, “Geometric features analysis for the classification of cultural heritage point clouds,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W15, pp. 541–548, 08 2019.
- [27] S. Yang, M. Hou, and S. Li, “Three-dimensional point cloud semantic segmentation for cultural heritage: A comprehensive review,” *Remote Sensing*, vol. 15, no. 3, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/3/548>
- [28] C. Tommasi, C. Achille, and F. Fassi, “From point cloud to bim: A modelling challenge in the cultural heritage field,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B5, pp. 429–436, 06 2016.
- [29] M. Baptista, “Documenting a complex modern heritage building using multi image close range photogrammetry and 3d laser scanned point clouds,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5/W2, pp. 675–678, 07 2013.
- [30] Z. Dong, B. Yang, F. Liang, R. Huang, and S. Scherer, “Hierarchical registration of unordered tls point clouds based on binary shape context descriptor,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 144, pp. 61–79, 2018.
- [31] S. Albrecht, “Transparent object reconstruction and registration confidence measures for 3d point clouds based on data inconsistency and viewpoint analysis,” 2018.

- [32] X. Huang, G. Mei, J. Zhang, and R. Abbas, "A comprehensive survey on point cloud registration," 2021.
- [33] M. M. Shanoer and F. M. Abed, "Evaluate 3d laser point clouds registration for cultural heritage documentation," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 21, no. 3, pp. 295–304, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1110982317300807>
- [34] L. Di Angelo, P. Di Stefano, A. Morabito, E. Guardiani *et al.*, "Point clouds registration based on constant radius features for large and detailed cultural heritage objects," in *2020 IMEKO TC-4 International Conference on Metrology for Archaeology and Cultural Heritage*. International Measurement Confederation (IMEKO), 2020, pp. 168–173.
- [35] E. Tournas and M. Tsakiri, "Automatic 3d point cloud registration for cultural heritage documentation," p. W8, 2009.
- [36] F. Tombari and F. Remondino, "Feature-based automatic 3d registration for cultural heritage applications," in *2013 Digital Heritage International Congress (DigitalHeritage)*, vol. 1. IEEE, 2013, pp. 55–62.
- [37] M. A. Butt and P. Maragos, "Optimum design of chamfer distance transforms," *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1477–1484, 1998.
- [38] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE International Conference on Robotics and Automation*, 2009, pp. 3212–3217.
- [39] P. BESL, "A method for registration of 3-d shapes," *Trans. PAMI*, vol. 14, no. 2, 1992.
- [40] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *International journal of computer vision*, vol. 13, no. 2, pp. 119–152, 1994.
- [41] Z. Zhang, Y. Dai, and J. Sun, "Deep learning based point cloud registration: an overview," *Virtual Reality & Intelligent Hardware*, vol. 2, no. 3, pp. 222–246, 2020, 3D Visual Processing and Reconstruction Special Issue. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2096579620300383>

- [42] K. Mirzaei, M. Arashpour, E. Asadi, H. Masoumi, Y. Bai, and A. Behnood, “3d point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review,” *Advanced Engineering Informatics*, vol. 51, p. 101501, 2022.
- [43] E. Agapaki and I. Brilakis, “Cloi-net: Class segmentation of industrial facilities’ point cloud datasets,” *Advanced Engineering Informatics*, vol. 45, p. 101121, 2020.
- [44] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [45] J. Markel, “The sift algorithm for fundamental frequency estimation,” *IEEE Transactions on Audio and Electroacoustics*, vol. 20, no. 5, pp. 367–377, 1972.
- [46] Z. Jiao, R. Liu, P. Yi, and D. Zhou, “A point cloud registration algorithm based on 3d-sift,” *Transactions on Edutainment XV*, pp. 24–31, 2019.
- [47] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” *Lecture notes in computer science*, vol. 3951, pp. 404–417, 2006.
- [48] M. P. Segundo, L. Gomes, O. R. P. Bellon, and L. Silva, “Automating 3d reconstruction pipeline by surf-based alignment,” in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 1761–1764.
- [49] Y. Zhong, “Intrinsic shape signatures: A shape descriptor for 3d object recognition,” in *2009 IEEE 12th international conference on computer vision workshops, ICCV workshops*. IEEE, 2009, pp. 689–696.
- [50] I. Sipiran and B. Bustos, “Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes,” *The Visual Computer*, vol. 27, pp. 963–976, 2011.
- [51] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, “Recognizing objects in range data using regional point descriptors,” in *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part III 8*. Springer, 2004, pp. 224–237.

- [52] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, “Aligning point cloud views using persistent feature histograms,” in *2008 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2008, pp. 3384–3391.
- [53] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, pp. 91–110, 2004.
- [54] F. Tombari, S. Salti, and L. Di Stefano, “Unique signatures of histograms for local surface description,” in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part III 11*. Springer, 2010, pp. 356–369.
- [55] H. Van Nguyen and F. Porikli, “Concentric ring signature (cors) for 3d object detection recognition, and registration,” *Elsevier Pattern Recognition*, 2011.
- [56] X. Guo, “Three dimensional moment invariants under rigid transformation,” in *Computer Analysis of Images and Patterns: 5th International Conference, CAIP’93 Budapest, Hungary, September 13–15, 1993 Proceedings 5*. Springer, 1993, pp. 518–522.
- [57] X. Li and I. Guskov, “Multiscale features for approximate alignment of point-based surfaces.” in *Symposium on geometry processing*, vol. 255, 2005, pp. 217–226.
- [58] X. Zhang, H. Li, Z. Cheng *et al.*, “Curvature estimation of 3d point cloud surfaces through the fitting of normal section curvatures,” *Proceedings of ASIAGRAPH*, vol. 2008, pp. 23–26, 2008.
- [59] A. Haar, *Zur theorie der orthogonalen funktionensysteme*. Georg-August-Universitat, Gottingen., 1909.
- [60] J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool, “Hough transform and 3d surf for robust three dimensional classification,” in *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part VI 11*. Springer, 2010, pp. 589–602.
- [61] P. Scovanner, S. Ali, and M. Shah, “A 3-dimensional sift descriptor and its application to action recognition,” in *Proceedings of the 15th ACM international conference on Multimedia*, 2007, pp. 357–360.

- [62] J. Li, Q. Hu, and M. Ai, “Rift: Multi-modal image matching based on radiation-invariant feature transform,” *arXiv preprint arXiv:1804.09493*, 2018.
- [63] R. B. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.
- [64] S. Lazebnik, C. Schmid, and J. Ponce, “A sparse texture representation using local affine regions,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1265–1278, 2005.
- [65] A. Ranganathan, “The levenberg-marquardt algorithm,” *Tutorial on LM algorithm*, vol. 11, no. 1, pp. 101–110, 2004.
- [66] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, and H. Voos, “Visual slam: What are the current trends and what to expect?” *Sensors*, vol. 22, no. 23, p. 9297, 2022.
- [67] F. Huang, H. Yang, X. Tan, S. Peng, J. Tao, and S. Peng, “Fast reconstruction of 3d point cloud model using visual slam on embedded uav development platform,” *Remote Sensing*, vol. 12, no. 20, p. 3308, 2020.
- [68] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann, “6d slam—3d mapping outdoor environments,” *Journal of Field Robotics*, vol. 24, no. 8-9, pp. 699–722, 2007.
- [69] S. Y. Loo, S. Mashohor, S. H. Tang, and H. Zhang, “Deeprelativefusion: Dense monocular slam using single-image relative depth prediction,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 6641–6648.
- [70] Y. Zhang, J. Yu, X. Huang, W. Zhou, and J. Hou, “Pcr-cg: Point cloud registration via deep explicit color and geometry,” in *European Conference on Computer Vision*. Springer, 2022, pp. 443–459.
- [71] O. Choi and W. Hwang, “Colored point cloud registration by depth filtering,” *Sensors*, vol. 21, no. 21, p. 7023, 2021.
- [72] Z. Hui, Z. Yong-Jian, Z. Lei, J. Xiao-Xue, and L. Li-Ying, “Fast color point cloud registration based on virtual viewpoint image,” *Frontiers in Physics*, vol. 10, p. 1021, 2022.

- [73] D. Bojanić, K. Bartol, T. Petković, and T. Pribanić, “A review of rigid 3d registration methods,” in *13th International Scientific-Professional Symposium Textile Science and Economy*, 2020, pp. 286–296.
- [74] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
- [75] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural networks*, vol. 61, pp. 85–117, 2015.
- [76] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, “3dmatch: Learning local geometric descriptors from rgb-d reconstructions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1802–1811.
- [77] P. Cunningham, M. Cord, and S. J. Delany, “Supervised learning,” *Machine learning techniques for multimedia: case studies on organization and retrieval*, pp. 21–49, 2008.
- [78] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [79] M. Khoury, Q.-Y. Zhou, and V. Koltun, “Learning compact geometric features,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [80] H. Deng, T. Birdal, and S. Ilic, “Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 602–618.
- [81] —, “Ppf-foldnet: Unsupervised learning of rotation invariant 3d local descriptors,” 09 2018, pp. 620–638.
- [82] Y. Yang, C. Feng, Y. Shen, and D. Tian, “Foldingnet: Point cloud auto-encoder via deep grid deformation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 206–215.

- [83] Z. J. Yew and G. H. Lee, “3dfeat-net: Weakly supervised local 3d features for point cloud registration,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 607–623.
- [84] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, “Signature verification using a”siamese”time delay neural network,” *Advances in neural information processing systems*, vol. 6, 1993.
- [85] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons, 2007.
- [86] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, “Pointnetlk: Robust & efficient point cloud registration using pointnet,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7163–7172.
- [87] X. Li, J. K. Pontes, and S. Lucey, “Pointnetlk revisited,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 763–12 772.
- [88] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *IJCAI’81: 7th international joint conference on Artificial intelligence*, vol. 2, 1981, pp. 674–679.
- [89] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “Pointnet++: Deep hierarchical feature learning on point sets in a metric space,” *Advances in neural information processing systems*, vol. 30, 2017.
- [90] H. Abdi, “Singular value decomposition (svd) and generalized singular value decomposition,” *Encyclopedia of measurement and statistics*, vol. 907, p. 912, 2007.
- [91] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from [tensorflow.org](https://www.tensorflow.org). [Online]. Available: <https://www.tensorflow.org/>

- [92] Y. Wang and J. M. Solomon, “Deep closest point: Learning representations for point cloud registration,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 3523–3532.
- [93] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, “Dynamic graph cnn for learning on point clouds,” *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.
- [94] A. Santoro, D. Raposo, D. G. Barrett, M. Malinowski, R. Pascanu, P. Battaglia, and T. Lillicrap, “A simple neural network module for relational reasoning,” *Advances in neural information processing systems*, vol. 30, 2017.
- [95] Y. Wang and J. M. Solomon, “Prnet: Self-supervised learning for partial-to-partial registration,” *Advances in neural information processing systems*, vol. 32, 2019.
- [96] J. C. Gower and G. B. Dijkstra, *Procrustes problems*. OUP Oxford, 2004, vol. 30.
- [97] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, “A survey on contrastive self-supervised learning,” *Technologies*, vol. 9, no. 1, p. 2, 2020.
- [98] V. Sarode, X. Li, H. Goforth, Y. Aoki, R. A. Srivatsan, S. Lucey, and H. Choset, “Pernet: Point cloud registration network using pointnet encoding,” *arXiv preprint arXiv:1908.07906*, 2019.
- [99] Z. Zhang, Y. Dai, and J. Sun, “Deep learning based point cloud registration: an overview,” *Virtual Reality & Intelligent Hardware*, vol. 2, no. 3, pp. 222–246, 2020.
- [100] X. Huang, G. Mei, and J. Zhang, “Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 366–11 374.
- [101] H. Deng, T. Birdal, and S. Ilic, “3d local features for direct pairwise registration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3244–3253.
- [102] C. Choy, J. Park, and V. Koltun, “Fully convolutional geometric features,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 8958–8966.

- [103] C. Choy, W. Dong, and V. Koltun, “Deep global registration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2514–2523.
- [104] Z. Gojcic, C. Zhou, J. D. Wegner, L. J. Guibas, and T. Birdal, “Learning multiview 3d point cloud registration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1759–1769.
- [105] E. Gallo, “The $so(3)$ and $se(3)$ lie algebras of rigid body rotations and motions and their application to discrete integration, gradient descent optimization, and state estimation,” *arXiv preprint arXiv:2205.12572*, 2022.
- [106] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, and C. Rother, “Dsac-differentiable ransac for camera localization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6684–6692.
- [107] E. Brachmann and C. Rother, “Learning less is more-6d camera localization via 3d surface regression,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4654–4662.
- [108] d. Björck and V. Pereyra, “Solution of vandermonde systems of equations,” *Mathematics of computation*, vol. 24, no. 112, pp. 893–903, 1970.
- [109] G. D. Pais, S. Ramalingam, V. M. Govindu, J. C. Nascimento, R. Chellappa, and P. Miraldo, “3dregnet: A deep neural network for 3d point registration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7193–7203.
- [110] D. Bojanić, K. Bartol, J. Forest, T. Petković, and T. Pribanić, “Deep learning vs. traditional 3d registration: A featureless 3d registration baseline,” *Traditional 3d Registration: A Featureless 3d Registration Baseline*.
- [111] X. Bai, Z. Luo, L. Zhou, H. Fu, L. Quan, and C.-L. Tai, “D3feat: Joint learning of dense detection and description of 3d local features,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6359–6367.
- [112] H. Thomas, C. R. Qi, J.-E. Deschard, B. Marcotegui, F. Goulette, and L. J. Guibas, “Kpconv: Flexible and deformable convolution for point clouds,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6411–6420.

- [113] J. Li and G. H. Lee, “Usip: Unsupervised stable interest point detection from 3d point clouds,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 361–370.
- [114] S. Huang, Z. Gojcic, M. Usvyatsov, A. Wieser, and K. Schindler, “Predator: Registration of 3d point clouds with low overlap,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 4267–4276.
- [115] “Predator: Registration of 3d point clouds with low overlap.” mIT License. [Online]. Available: <https://github.com/prs-eth/OverlapPredator>
- [116] S. Ao, Q. Hu, B. Yang, A. Markham, and Y. Guo, “Spinnet: Learning a general surface descriptor for 3d point cloud registration,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 11 753–11 762.
- [117] F. Poiesi and D. Boscaini, “Learning general and distinctive 3d local deep descriptors for point cloud registration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [118] J. Yang, Q. Zhang, Y. Xiao, and Z. Cao, “Toldi: An effective and robust approach for 3d local shape description,” *Pattern Recognition*, vol. 65, pp. 175–187, 2017.
- [119] W. Yuan, D. Held, C. Mertz, and M. Hebert, “Iterative transformer network for 3d point cloud,” *arXiv preprint arXiv:1811.11209*, 2018.
- [120] M. Khoury, Q.-Y. Zhou, and V. Koltun, “Learning compact geometric features,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 153–161.
- [121] F. Poiesi and D. Boscaini, “Distinctive 3d local deep descriptors,” in *2020 25th International conference on pattern recognition (ICPR)*. IEEE, 2021, pp. 5720–5727.
- [122] H. Wang, Y. Liu, Z. Dong, and W. Wang, “You only hypothesize once: Point cloud registration with rotation-equivariant descriptors,” in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 1630–1641.
- [123] L. Li, S. Zhu, H. Fu, P. Tan, and C.-L. Tai, “End-to-end learning local multi-view descriptors for 3d point clouds,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1919–1928.

- [124] S. Liu, T. Li, W. Chen, and H. Li, “Soft rasterizer: A differentiable renderer for image-based 3d reasoning,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7708–7717.
- [125] X. Huang, W. Qu, Y. Zuo, Y. Fang, and X. Zhao, “Imfnet: Interpretable multimodal fusion for point cloud registration,” *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 323–12 330, 2022.
- [126] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [127] J. Lee, S. Kim, M. Cho, and J. Park, “Deep hough voting for robust global registration,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 994–16 003.
- [128] W. Zhou, C. Ma, and A. Kuijper, “Hough-space-based hypothesis generation and hypothesis verification for 3d object recognition and 6d pose estimation,” *Computers & Graphics*, vol. 72, pp. 122–134, 2018.
- [129] X. Bai, Z. Luo, L. Zhou, H. Chen, L. Li, Z. Hu, H. Fu, and C.-L. Tai, “Pointdsc: Robust point cloud registration using deep spatial consistency,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15 859–15 869.
- [130] M. Leordeanu and M. Hebert, “A spectral technique for correspondence problems using pairwise constraints,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, vol. 2. IEEE, 2005, pp. 1482–1489.
- [131] J. Wang, C. Yang, L. Wei, and R. Chen, “Csce-net: Channel-spatial contextual enhancement network for robust point cloud registration,” *Remote Sensing*, vol. 14, no. 22, p. 5751, 2022.
- [132] H. Yang and L. Carlone, “A polynomial-time solution for robust registration with extreme outlier rates,” *arXiv preprint arXiv:1903.08588*, 2019.
- [133] H. Yang, J. Shi, and L. Carlone, “Teaser: Fast and certifiable point cloud registration,” *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 314–333, 2020.

- [134] X. Huang, Z. Liang, C. Bajaj, and Q. Huang, "Translation synchronization via truncated least squares," *Advances in neural information processing systems*, vol. 30, 2017.
- [135] Z. Zhang, J. Sun, Y. Dai, B. Fan, and M. He, "Vrnet: Learning the rectified virtual corresponding points for 3d point cloud registration," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 4997–5010, 2022.
- [136] W. Chen, H. Li, Q. Nie, and Y.-H. Liu, "Deterministic point cloud registration via novel transformation decomposition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6348–6356.
- [137] Z. Chen, K. Sun, F. Yang, and W. Tao, "Sc2-pcr: A second order spatial compatibility for efficient and robust point cloud registration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 13 221–13 231.
- [138] Z. J. Yew and G. H. Lee, "Rpm-net: Robust point matching using learned features," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 824–11 833.
- [139] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *The annals of mathematical statistics*, vol. 35, no. 2, pp. 876–879, 1964.
- [140] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [141] K. Fu, S. Liu, X. Luo, and M. Wang, "Robust point cloud registration framework based on deep graph matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8893–8902.
- [142] K. Fischer, M. Simon, F. Olsner, S. Milz, H.-M. Gross, and P. Mader, "Stickypillars: Robust and efficient feature matching on point clouds using graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 313–323.
- [143] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.

- [144] J. Li, C. Zhang, Z. Xu, H. Zhou, and C. Zhang, “Iterative distance-aware similarity matrix convolution with mutual-supervised point elimination for efficient point cloud registration,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*. Springer, 2020, pp. 378–394.
- [145] A.-Q. Cao, G. Puy, A. Boulch, and R. Marlet, “Pcam: Product of cross-attention matrices for rigid registration of point clouds,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 229–13 238.
- [146] W. Yuan, B. Eckart, K. Kim, V. Jampani, D. Fox, and J. Kautz, “Deepgmr: Learning latent gaussian mixture models for registration,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 2020, pp. 733–750.
- [147] D. A. Reynolds *et al.*, “Gaussian mixture models.” *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.
- [148] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [149] H. Xu, S. Liu, G. Wang, G. Liu, and B. Zeng, “Omnet: Learning overlapping mask for partial-to-partial point cloud registration,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3132–3141.
- [150] Z. J. Yew and G. H. Lee, “Regtr: End-to-end point cloud correspondences with transformers,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6677–6686.
- [151] Z. Qin, H. Yu, C. Wang, Y. Guo, Y. Peng, and K. Xu, “Geometric transformer for fast and robust point cloud registration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 11 143–11 152.
- [152] J. Xu, Y. Huang, Z. Wan, and J. Wei, “Glorn: Strong generalization fully convolutional network for low-overlap point cloud registration,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

- [153] Y. Li and T. Harada, “Leopard: Learning partial point cloud matching in rigid and deformable scenes,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5554–5564.
- [154] X. Bai, “awesome-point-cloud-registration.” [Online]. Available: <https://github.com/XuyangBai/awesome-point-cloud-registration>
- [155] D. Aiger, N. J. Mitra, and D. Cohen-Or, “4-points congruent sets for robust pairwise surface registration,” in *ACM SIGGRAPH 2008 papers*, 2008, pp. 1–10.
- [156] N. Mellado, D. Aiger, and N. J. Mitra, “Super 4pcs fast global pointcloud registration via smart indexing,” in *Computer graphics forum*, vol. 33, no. 5. Wiley Online Library, 2014, pp. 205–215.
- [157] V. Sarode, A. Dhagat, R. A. Srivatsan, N. Zevallos, S. Lucey, and H. Choset, “Masknet: A fully-convolutional network to estimate inlier points,” in *2020 International Conference on 3D Vision (3DV)*. IEEE, 2020, pp. 1029–1038.
- [158] K. Lee, J. Lee, and J. Park, “Learning to register unbalanced point pairs,” *arXiv preprint arXiv:2207.04221*, 2022.
- [159] H. Chen, Z. Wei, Y. Xu, M. Wei, and J. Wang, “Imlovenet: Misaligned image-supported registration network for low-overlap point cloud pairs,” in *ACM SIGGRAPH 2022 Conference Proceedings*, 2022, pp. 1–9.
- [160] Z. Chen, F. Yang, and W. Tao, “Detarnet: Decoupling translation and rotation by siamese network for point cloud registration,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 401–409.
- [161] H. Lim, S. Yeon, S. Ryu, Y. Lee, Y. Kim, J. Yun, E. Jung, D. Lee, and H. Myung, “A single correspondence is enough: Robust global registration to avoid degeneracy in urban environments,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 8010–8017.
- [162] Z. Chen, H. Chen, L. Gong, X. Yan, J. Wang, Y. Guo, J. Qin, and M. Wei, “Utopic: Uncertainty-aware overlap prediction network for partial point cloud registration,” *arXiv preprint arXiv:2208.02712*, 2022.

- [163] W. Liu, H. Wu, and G. S. Chirikjian, “Lsg-cpd: Coherent point drift with local surface geometry for point cloud registration,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 293–15 302.
- [164] A. Myronenko and X. Song, “Point set registration: Coherent point drift,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [165] I. J. Myung, “Tutorial on maximum likelihood estimation,” *Journal of mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.
- [166] X.-L. Meng and D. Van Dyk, “The em algorithm—an old folk-song sung to a fast new tune,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 59, no. 3, pp. 511–567, 1997.
- [167] N. Lang and J. M. Francos, “Deepume: Learning the universal manifold embedding for robust point cloud registration,” *arXiv preprint arXiv:2112.09938*, 2021.
- [168] A. Efraim and J. M. Francos, “The universal manifold embedding for estimating rigid transformations of point clouds,” in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 5157–5161.
- [169] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
- [170] R. Stojnic, R. Taylor, M. Kardas, E. Saravia, G. Cucurull, T. Scialom, and M. AI, “Papers with code,” cC-BY-SA License. [Online]. Available: <https://paperswithcode.com/>
- [171] S. Horache, J.-E. Deschaud, and F. Goulette, “3d point cloud registration with multi-scale architecture and unsupervised transfer learning,” in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 1351–1361.
- [172] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

- [173] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, “Searching efficient 3d architectures with sparse point-voxel convolution,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII*. Springer, 2020, pp. 685–702.
- [174] L. Torrey and J. Shavlik, “Transfer learning,” in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.
- [175] Z. Gojcic, C. Zhou, J. D. Wegner, and A. Wieser, “The perfect match: 3d point cloud matching with smoothed densities,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5545–5554.
- [176] X. Cheng, S. Yan, Y. Liu, M. Zhang, and C. Chen, “R-pcr: Recurrent point cloud registration using high-order markov decision,” *Remote Sensing*, vol. 15, no. 7, p. 1889, 2023.
- [177] W.-K. Ching, X. Huang, M. K. Ng, T.-K. Siu, W.-K. Ching, X. Huang, M. K. Ng, and T.-K. Siu, “Higher-order markov chains,” *Markov Chains: Models, Algorithms and Applications*, pp. 141–176, 2013.
- [178] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” *arXiv preprint arXiv:1409.1259*, 2014.
- [179] L. Zhu, H. Guan, C. Lin, and R. Han, “Leveraging inlier correspondences proportion for point cloud registration,” *arXiv preprint arXiv:2201.12094*, 2022.
- [180] A. Zeng, “3dmatch: Learning local geometric descriptors from rgb-d reconstructions.” mIT License. [Online]. Available: <https://3dmatch.cs.princeton.edu/#rgb-d-reconstruction-datasets>
- [181] H. Yu, F. Li, M. Saleh, B. Busam, and S. Ilic, “Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 23 872–23 884, 2021.
- [182] W. Holnicki, “The king’s chinese cabinet,” 2012. [Online]. Available: https://gis.muzeum-wilanow.pl/files/zalaczniki_pom/C50490.jpg

- [183] “The king’s wardrobe.” [Online]. Available: https://gis.muzeum-wilanow.pl/files/zalaczniki_pom/C79282.jpg
- [184] Z. Reszka, “Queen’s antecabinet,” 2011. [Online]. Available: https://gis.muzeum-wilanow.pl/files/zalaczniki_pom/C42540.jpg
- [185] A. Indyk, “Apollo i sybilla kumejska,” 2008. [Online]. Available: https://gis.muzeum-wilanow.pl/files/zalaczniki_pom/C13896.jpg
- [186] “Geographic information system museum of king jan iii’s palace at wilanow.” [Online]. Available: <https://gis.muzeum-wilanow.pl/gis/eng/>
- [187] R. Sitnik, E. Bunsch, G. Maczkowski, W. Załuski, K. Lech, J. Michoński, J. Krzesłowski, and P. Foryś, “Towards automated, high resolution 3d scanning of large surfaces for cultural heritage documentation,” *Electronic Imaging*, vol. 2016, no. 21, pp. 3DIPM–051, 2016.
- [188] P. Foryś, K. Lech, J. Michoński, Ł. Markiewicz, E. Bunsch, and R. Sitnik, “Semi-automated method for views integration based on alternative color and geometry characteristics,” in *Optical Sensors 2019*, vol. 11028. SPIE, 2019, pp. 359–368.
- [189] A. Bundy and L. Wallen, “Difference of gaussians,” *Catalogue of Artificial Intelligence Tools*, pp. 30–30, 1984.
- [190] P. Foryś, R. Sitnik, J. Markiewicz, and E. Bunsch, “Fast adaptive multimodal feature registration (FAMFR) - an effective high resolution point clouds registration workflow for cultural heritage interiors,” *Heritage Science (W RECENZJI)*.
- [191] P. Liberadzki, Ł. Markiewicz, M. Witkowski, and R. Sitnik, “Novel 4d whole body scanning solution and its medical application,” in *Proceedings of the 9th International Conference and Exhibition on 3D Body Scanning and Processing Technologies*, 2018, pp. 47–56.
- [192] R. I.-R. BT *et al.*, “Studio encoding parameters of digital television for standard 4: 3 and wide-screen 16: 9 aspect ratios,” *International Radio Consultative Committee International Telecommunication Union, Switzerland, CCIR Rep*, 2011.

- [193] A. S. Householder, *The Numerical Treatment of a Single Nonlinear Equation 1970*. McGraw Hill.
- [194] G. Sidorov, A. Gelbukh, H. Gómez-Adorno, and D. Pinto, “Soft similarity and soft cosine measure: Similarity of features in vector space model,” *Computación y Sistemas*, vol. 18, no. 3, pp. 491–504, 2014.
- [195] S. Umeyama, “Least-squares estimation of transformation parameters between two point patterns,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 04, pp. 376–380, 1991.
- [196] “Learning3d: a modern library for deep learning on 3d point clouds data.” MIT License. [Online]. Available: <https://github.com/vinits5/learning3d>
- [197] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao *et al.*, “A deep representation for volumetric shapes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, 2015, pp. 1912–1920.
- [198] “Geometric transformer for fast and robust point cloud registration.” MIT License. [Online]. Available: <https://github.com/qinzheng93/GeoTransformer>
- [199] K. Chowdhary and K. Chowdhary, “Natural language processing,” *Fundamentals of artificial intelligence*, pp. 603–649, 2020.
- [200] D. Zhang, J. Wang, and X. Zhao, “Estimating the uncertainty of average f1 scores,” in *Proceedings of the 2015 International conference on the theory of information retrieval*, 2015, pp. 317–320.
- [201] J. Rutkiewicz, M. Malesa, M. Karaszewski, P. Foryś, P. Siekański, and R. Sitnik, “The method of acquiring and processing 3d data from drones,” in *Speckle 2018: VII International Conference on Speckle Metrology*, vol. 10834. SPIE, 2018, pp. 558–564.
- [202] “Mnemosiis,” Apr 2023. [Online]. Available: <https://mnemosiis.pl/>