



Łódź, 8-10 kwietnia 2015

OPTYCZNY SKANER OBIEKTÓW TRÓJWYMIAROWYCH

Streszczenie: W pracy przedstawiono technikę skanowania obiektów trójwymiarowych. Model 3D obiektu tworzony jest jedynie na podstawie analizy obrazów zarejestrowanych za pomocą zwykłej kamery (wideo). Zaproponowana technika wykorzystuje, opracowywany w ramach Międzynarodowej Organizacji Normalizacyjnej (ISO), algorytm estymacji głębi stereoskopowej z szeregu niezrektyfikowanych obrazów. Efektywność opracowanego rozwiązania sprawdzono na wielu obiektach testowych, zarejestrowanych za pomocą sensora Kinect. W celach porównawczych, poza obrazem obiektu, zarejestrowano również mapę głębi. Przedstawione wyniki eksperymentów potwierdzają możliwość zastąpienia urządzeń aktywnych do tworzenia trójwymiarowych modeli obiektów urządzeniami pasywnymi, wykorzystującymi techniki algorytmicznej estymacji map głębi jedynie na podstawie analizy zarejestrowanego obrazu.

1. WSTĘP

Wraz z gwałtownym rozwojem dziedziny trójwymiarowej grafiki komputerowej, filmowych efektów specjalnych, przemysłu gier komputerowych oraz wchodzących na rynek drukarek 3D, można zaobserwować rosnące zainteresowanie rozwiązaniami pozwalającymi na szybkie i dokładne tworzenie modeli 3D fizycznych przedmiotów. Ręczne tworzenie modeli jest bardzo czasochłonne i wymaga specyficznych umiejętności oraz pewnej dozy talentu. Z kolei automatyczne techniki tworzenia modeli 3D, zwane skanerami 3D, wymagają specjalistycznego sprzętu.

Na rynku są już dostępne różne systemy skanowania obiektów trójwymiarowych (3D), spośród których warto wymienić: kamery głębi [1], skanery wykorzystujące światło strukturalne w zakresie widzialnym lub w podczerwieni [2] oraz laserowe skanery obiektów. Jednak techniki te albo są bardzo drogie albo wytworzone z ich wykorzystaniem modele są niskiej rozdzielczości i jakości.

W takich okolicznościach, prosta i tania technika automatycznego tworzenia modeli przestrzennych wydaje się być wielce pożądana. Co więcej, gdyby takie modele udało się stworzyć, wykorzystując tylko powszechnie dostępne kamery wideo, można by zdecydowanie poszerzyć grono potencjalnych użytkowników.

Obecnie aktywnie rozwijane są techniki przetwarzania obrazu trójwymiarowego, a w szczególności techniki wyznaczania głębi stereoskopowej na podstawie dwóch lub więcej obrazów sceny. Techniki znane z literatury [3] zazwyczaj wymagają obrazów z pary kamer z równoległymi osiami optycznymi. Jednak ostatnio, w ramach prac

standaryzacyjnych, prowadzonych przez grupę ekspertów MPEG Międzynarodowej Organizacji Normalizacyjnej, nad telewizją swobodnego punktu widzenia, autorzy opracowali technikę estymacji map głębi pozwalającą na wyznaczanie głębi stereoskopowej z trzech niezrektyfikowanych obrazów [4]. Oprogramowanie to, tworzone we współpracy z ekspertami z całego świata, jest obecnie techniką odniesienia w pracach nad modelowaniem przestrzennym sceny ruchomej.

Autorzy niniejszej pracy postawili sobie pytanie czy, wykorzystując ostatnie postępy z zakresu technik wyznaczania głębi stereoskopowej, możliwe jest stworzenie prostej i taniej techniki skanowania obiektów 3D o wystarczającej jakości i wysokiej rozdzielczości.

2. CEL I ZAŁOŻENIA

Opracowana technika korzysta jedynie z obrazów skanowanego obiektu, zarejestrowanych przez kamerę podczas obchodzenia obiektu przez operatora. Na podstawie analizy obrazu wyznaczana jest mapa głębi dla każdego obrazu, a następnie, w pełni automatycznie, konstruowany jest model 3D skanowanego obiektu. Użytkany w ten sposób model może zostać zapisany w jednym z popularnych formatów graficznych, wykorzystywanych w wiodących programach do obróbki grafiki trójwymiarowej, np. 3D Studio MAX, Blender. Model może zostać użyty wprost, na przykład może zostać wydrukowany za pomocą drukarki 3D lub może stanowić punkt wejściowy dla dalszego przetwarzania i edycji.

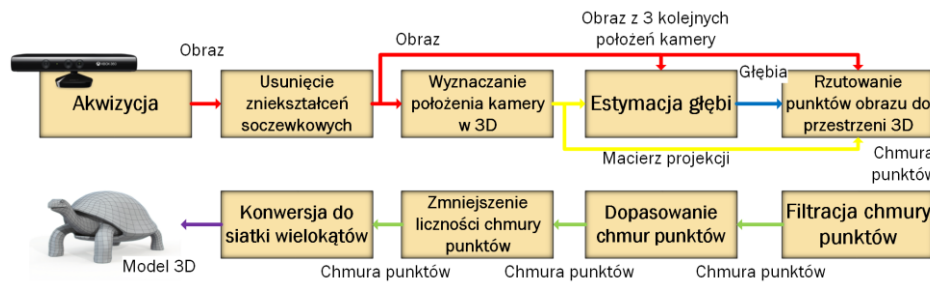
W kolejnych punktach przedstawiono szczegółowy opis systemu oraz wyniki przeprowadzonych eksperymentów, mających na celu ocenę systemu.

3. OPIS SYSTEMU

Opracowany system do skanowania obiektów korzysta jedynie z obrazu z kamery obserwującej skanowany obiekt, kolejno ze wszystkich stron. Na tej podstawie rekonstruowana jest informacja przestrzenna niezbędna do stworzenia modelu 3D skanowanego obiektu. Schemat systemu przedstawiono na rysunku 1.

3.1. Akwizycja obrazu

Akwizycja obrazu odbywa się przy pomocy kamery trzymanej w ręku operatora, rejestrującej obraz skanowanego obiektu. Operator ma za zadanie obejść obiekt tak, aby zarejestrować go ze wszystkich stron.



Rys. 1 Schemat przedstawiający opracowany algorytm otrzymywania modeli trójwymiarowych wykorzystujący analizę obrazu i estymację głębi stereoskopowej.

Wykorzystywaną kamerę należy uprzednio skalibrować w celu wyznaczenia jej parametrów wewnętrznych (ogniskowej oraz współrzędnych punktu głównego), a także w celu wyznaczenia i usunięcia zniekształceń soczewkowych zastosowanej kamery. Istnieje wiele metod kalibracji kamery. W pracy wykorzystano najpopularniejszą metodę – metodę Zhanga [5].

W pracy wykorzystano sensor Kinect X360. Pełnił on rolę kamery rejestrującej obraz obiektu o rozdzielczości 640x480 punktów oraz, jedynie w celach porównawczych, informację przestrzenną w postaci mapy głębi o rozdzielczości 320x240 punktów. Zapisanie informacji o głębi rejestrowanego obrazu pozwoliło nam na porównanie dokładności modeli, stworzonych na podstawie algorytmicznej estymacji głębi, z tymi, stworzonymi na podstawie zarejestrowanych map głębi.

3.2. Lokalizacja kamery

Na podstawie analizy każdego ujęcia należy wyznaczyć położenie kamery, z którego obraz został zarejestrowany, względem skanowanego obiektu, w jednym, globalnym układzie współrzędnych. W literaturze znanych jest wiele metod, pozwalających na wyznaczenie położenia kamery na podstawie i względem rejestrowanego obrazu [6]. Zasadniczo metody te można podzielić na dwa typy: metody wykorzystujące znaczniki umieszczone w scenie oraz metody nie wykorzystujące takich znaczników. Skuteczność metod nie wykorzystujących znaczników w scenie silnie zależy od treści analizowanych/ rejestrowanych obrazów. Główny problem związany jest z wyznaczaniem punktów charakterystycznych w obrazach, podczas obserwacji obiektu z różnych stron oraz późniejszym dopasowaniem ich do siebie (zidentyfikowanie tych samych punktów w różnych obrazach).

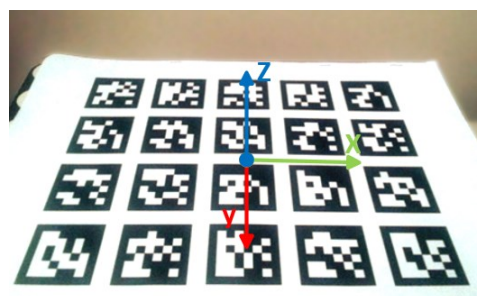
W związku z powyższym, w pracy zastosowano prostszą, ale dającą lepsze rezultaty metodę, opartą o znacznik umieszczony w scenie [7] (rys. 2). Znacznik został tak dobrany, aby umożliwić proste oraz bardzo dokładne ustalenie położenia kamery nawet wtedy, gdy część znacznika będzie przesłonięta przez skanowany obiekt.

Dla każdego ujęcia skanowanego obiektu wyznaczone jest położenie kamery w przestrzeni względem układu współrzędnych, związanego ze znacznikiem (rys. 2).

3.1. Wyznaczanie informacji o głębi

W wyniku procesu rejestracji obrazu przez kamerę, czyli rzutowania punktów z przestrzeni 3D na płaszczy-

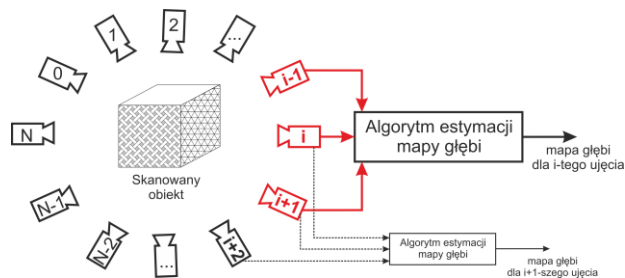
znę obrazu, tracona jest informacja o ich odległości od kamery. Aby wyznaczyć położenie punktu w przestrzeni trójwymiarowej i w dalszej perspektywie stworzyć model 3D, niezbędne jest odzyskanie informacji o odległości punktów widzianych przez kamerę.



Rys. 2 Opracowany szablon z wieloma znacznikami z pokazanym globalnym układem współrzędnych rejestrowanej sceny.

Informację o głębi można uzyskać na dwa sposoby: wyznaczając ją na podstawie analizy obrazu sceny, przy różnych położeniach kamery (analizując różne ujęcia) lub zarejestrować ją za pomocą specjalnej kamery głębi. W pracy skupiono się na pierwszej metodzie, druga natomiast została użyta jedynie w celu weryfikacji poprawności procesu estymacji algorytmicznej map głębi.

W celu wyznaczenia głębi zastosowano, rozwijaną w ramach grupy MPEG Międzynarodowej Organizacji Normalizacyjnej (ISO), implementację algorytmu estymacji map głębi DERS (ang. Depth Estimation Reference Software) [4]. Algorytm analizuje jednocześnie obrazy z trzech kamer (w naszym przypadku trzech ujęć) i wyznacza mapę głębi dla wybranej (najczęściej środkowej) kamery (w naszym przypadku środkowego ujęcia). W ten sposób wyznacza się mapę głębi dla każdego położenia kamery (ujęcia), za każdym razem korzystając z 2 najbliższych (sąsiednich) położen kamery (ujęć) (rys. 3).



Rys. 3 Schemat wyznaczania mapy głębi dla każdego ujęcia skanowanego obiektu.

W wyniku, otrzymujemy zbiór obrazów skanowanego obiektu wraz z odpowiadającymi im mapami głębi.

3.2. Tworzenie modelu trójwymiarowego

Korzystając z zarejestrowanych obrazów, wyznaczonej informacji o głębi (map głębi) oraz informacji o położeniu kamery podczas rejestracji każdego obrazu (ujęcia), konstruuje się model 3D skanowanego obiektu. Dla każdego punktu każdego zarejestrowanego obrazu, na podstawie informacji o położeniu kamery w przestrzeni i wyznaczonej mapy głębi, wyznaczane jest położenie tego punktu w przestrzeni trójwymiarowej [8]:

$$M = z P^{-1}m$$

gdzie $m = [x, y, 1]^T$ jest wektorem opisującym położenie transformowanego punktu w obrazie, P jest macierzą projekcji kamery dokonującej rejestracji obrazu danego ujęcia i tworzona jest na podstawie wyznaczonego położenia kamery w przestrzeni i parametrów wewnętrznych użytej kamery (ogniskowa, położenie punktu centralnego), a z jest wyznaczoną odległością rekonstruowanego punktu od kamery, wyznaczoną za pomocą algorytmu estymacji głębi.

W ten sposób, dla każdego ustawienia kamery (ujęcia), otrzymujemy częściowy model skanowanego obiektu w postaci chmury punktów w przestrzeni 3D.

3.2.1. Dopasowanie chmur punktów

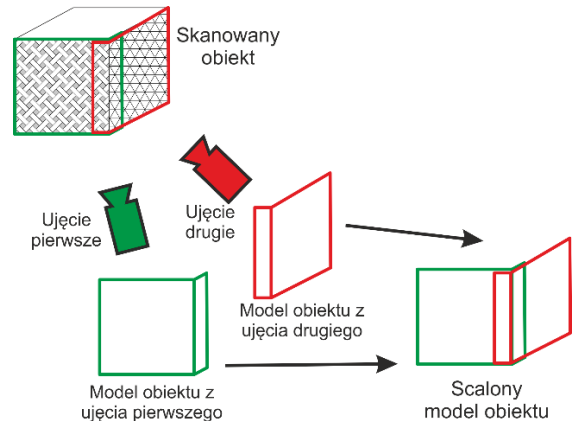
Aby stworzyć spójny model trójwymiarowy skanowanego obiektu, każda powstająca dla pojedynczego obrazu (ujęcia) chmura punktów musi zostać dopasowana do pozostałych chmur (rys. 4). Chmury punktów z różnych ujęć są podobne, ale mogą różnić się od siebie na skutek błędnego lub niedokładnego wyznaczenia położenia kamery lub błędów w wyznaczonej mapie głębi (odległości punktów od kamery). Również zakres pokrycia obiektu przez każdą indywidualną chmurę punktów jest nieco inny. W celu stworzenia w pełni spójnego modelu w postaci chmury punktów, kolejne chmury należy dopasować i połączyć w jeden model. Do tego celu wykorzystano algorytm ICP (Iterative Closest Point) [9]. Zadaniem algorytmu ICP jest wyznaczanie takiego przekształcenia (w postaci translacji i rotacji), aby zminimalizować średnią odległości euklidesowych pomiędzy (odpowiadającymi sobie) punktami dwóch chmur punktów.

W opracowanym rozwiązaniu, kolejne cząstkowe modele, w postaci chmur punktów, dopasowywane były kolejno do wynikowego modelu, w postaci chmury punktów, składającego się ze wszystkich poprzednio dopasowanych chmur punktów.

3.2.2. Uproszczenie modelu

Uzyskany model, w postaci chmury punktów, jest bardzo liczny. Po połączeniu indywidualnych modeli w jeden spójny model skanowanego obiektu, ten sam punkt w przestrzeni, na skutek rekonstrukcji z wielu ujęć, reprezentowany jest niepotrzebnie wiele razy w modelu. Ponadto uzyskany model zawiera pewną liczbę punktów, które nie reprezentują skanowanego obiektu, a są wynikiem szumu, błędów w estymacji głębi lub innych zakłóceń w procesie skanowania obiektu. Punkty takie znajdują się w pewnej odległości od głównej masy punktów, sta-

nowiących model skanowanego obiektu. Dlatego wynikiowy model jest filtrowany w celu usunięcia takich „odstających” punktów. Do filtracji chmury punktów, w celu usunięcia wartości tzw. „odstających”, wykorzystano algorytm ROR (Radius Outlier Removal) [10]. Zadaniem algorytmu jest usunięcie punktów, które w zadanej odległości mają liczbę sąsiadów mniejszą niż wartość zadanego parametrem algorytmu.



Rys. 4 Scalenie modeli obiektów powstałych z analizy pojedynczych ujęć.

Następnie zastosowano filtr mający na celu usunięcie powtarzających się punktów modelu, zmniejszenie liczności chmury punktów oraz jej gęstości. W tym celu zastosowano technikę polegającą na przyłożeniu do wyznaczonej chmury siatki, złożonej z niewielkich, sześciennych fragmentów, tworzących tzw. siatkę wokseli. Wszystkie punkty w obrębie jednego woksela zastępowane są jednym punktem o kolorze, będącym średnim kolorem wszystkich zastępowanych punktów oraz o położeniu, będącym centroidem położenia zastępowanych punktów. W ten sposób otrzymywany jest uproszczony model obiektu o zadanej gęstości.

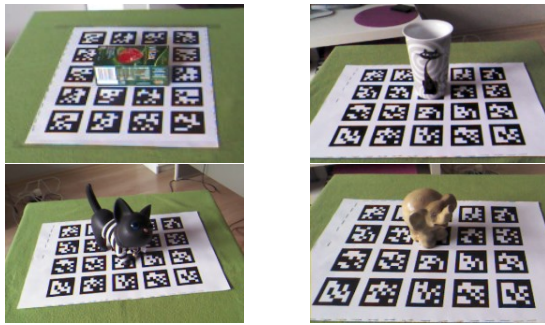
3.2.3. Transformacja do siatki wielokątów

Posiadając model obiektu, w postaci chmury punktów o odpowiedniej liczności/gęstości, możliwe jest przejście do ostatniego etapu automatycznego modelowania – do transformacji modelu w postaci chmury punktów do modelu zbudowanego z siatki wielokątów.

W celu transformacji chmury punktów do siatki wielokątów wykorzystano powszechnie dostępną implementację algorytmu Greedy Projection Triangulation [11]. W ten sposób uzyskano model, w postaci siatki wielokątów, który można łatwo przetwarzać w popularnych programach do obróbki grafiki komputerowej.

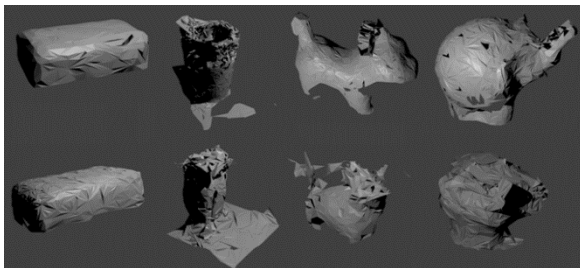
4. WYNIKI EKSPERYMENTALNE

Podczas pracy nad omawianą techniką zarejestrowano wiele sekwencji testowych. W niniejszej pracy przedstawione zostały przykładowe cztery: pudełko po herbacie, porcelanowy kubek oraz figurki kota i słonia (rys. 5). Sekwencje te różnią się stopniem skomplikowania skanowanego obiektu. Od prostego, regularnego pudełka po herbacie, poprzez trudniejszy kubek, aż do skomplikowanych kształtów figurek kota i słonia.



Rys. 5 Przykładowe ujęcia zarejestrowanych obiektów testowych.

Dla każdego skanowanego obiektu, na podstawie zarejestrowanych 40 ujęć (obrazów), zbudowano jego trójwymiarowy model, z wykorzystaniem opracowanej techniki. Dodatkowo, w celach porównawczych, zbudowano modele trójwymiarowe skanowanych obiektów, z wykorzystaniem zarejestrowanych bezpośrednio map głębi. Porównanie uzyskanych modeli przedstawiono na rysunku 6.



Rys. 6 Porównanie modeli uzyskanych jedynie na podstawie analizy obrazu (dolny wiersz) z modelami uzyskanymi z wykorzystaniem zarejestrowanych map głębi (górny wiersz).

Opracowany system, tworzący modele z wykorzystaniem algorytmicznej estymacji głębi, daje dobre rezultaty. Mimo, iż zawierają one ubytki i szum, prawidłowo odwzorowują proporcje, rozmiary modelowanych obiektów oraz ich ogólny kształt. Proste, regularne kształty, takie jak pudełko po herbacie czy kubek, modelowane są prawie tak idealnie, jak w przypadku wykorzystania map głębi zarejestrowanych za pomocą kamery głębi. Powoduje to, iż w kontekście modelowania trójwymiarowego należy pozytywnie ocenić przydatność algorytmu DERS, estymującego głębię jedynie na podstawie zarejestrowanych obrazów. Technika ta wydaje się być perspektywiczna, wymaga jednak dalszych badań ukierunkowanych na zwiększenie dokładności tworzonych modeli poprzez zwiększenie rozdzielczości rejestrowanych obrazów, zwiększenie dokładności wyznaczania położenia kamery, czy też zwiększenie skuteczności dopasowywania chmur punktów.

5. PODSUMOWANIE

Celem niniejszej pracy było opracowanie i implementacja algorytmu tworzącego modele trójwymiarowe, wyłącznie na podstawie analizy obrazów, rejestrowanych przez kamerę obserwującą skanowany obiekt. W tym celu, brakującą informację o głębi obrazu wyznaczono algorytmicznie, na podstawie obrazów pochodzących z 3 kolejnych położen kamery. W celach porównawczych

stworzono system skanowania, wspomagany informacją o głębi rejestrowaną przez sprzętowy sensor. Opracowany i zaimplementowany system tworzy modele trójwymiarowe, bazując tylko na obrazie z kamery, a wyniki skanowania są zapisywane do pliku w popularnym formacie graficznym. Umożliwia to edytowanie i przetwarzanie we wszystkich wiodących programach do obróbki grafiki 3D. Modele te mogą być wykorzystywane przez grafików 3D, projektantów wnętrz lub być wydrukowane za pomocą drukarki 3D. Ponadto, system wykorzystuje wyłącznie powszechnie dostępne elementy, jakimi są kamera rejestrująca obraz barwny oraz drukowane znaczniki. Zastosowanie takich prostych rozwiązań może ułatwić przyszłą komercjalizację przedstawionego rozwiązania.

PODZIĘKOWANIA

Praca finansowana ze środków przyznanych przez Ministerstwo Nauki i Szkolnictwa Wyższego na działalność statutową polegającą na prowadzeniu badań naukowych lub prac rozwojowych oraz zadań z nimi związanych, służących rozwojowi młodych naukowców oraz uczestników studiów doktoranckich.

SPIS LITERATURY

- [1] O. Choi, R. Horaud, "Time-of-Flight Cameras—Principles, Methods and Applications", Springer 2013
- [2] D. Scharstein, R. Szeliski, "High-accuracy stereo depth maps using structured light", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, str. 195-202, 2003
- [3] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International Journal of Computer Vision, nr 47, vol. 1-3, str. 7-42, 2002
- [4] K. Wegner, O. Stankiewicz, M. Tanimoto, M. Domanski, „Enhanced Depth Estimation Reference Software (DERS) for Free-viewpoint Television”, ISO/IEC JTC1/SC29/WG11 MPEG2013/M31518, Genewa, Szwajcaria, 2013
- [5] Z. Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, nr 11, str. 1330-1334, 2000
- [6] B. Cyganek, "Object Detection and Recognition in Digital Images: Theory and Practice", Wiley, 2013
- [7] V. Lepetit, P. Fua, "Monocular Model-Based 3D Tracking of Rigid Objects: A Survey, Foundations and Trends", Computer Graphics and Vision, nr 1, vol. 1, str. 1-89, 2005
- [8] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press 2003
- [9] S. Rusinkiewicz, M. Levoy, "Efficient variants of the ICP Algorithm", International Conference on 3-D Digital Imaging - 3DIM, str. 145-152, 2001
- [10] Point Cloud Library (PCL) – Documentation, <http://dev.pointclouds.org/projects/pcl/wiki>
- [11] Z.C. Marton, R.B. Rusu, M. Beetz, "On Fast Surface Reconstruction Methods for Large and Noisy Point Clouds", IEEE International Conference on Robotics and Automation, str. 3218-3223, 2009