XXIII Seminarium

ZASTOSOWANIE KOMPUTERÓW W NAUCE I TECHNICE' 2013
Oddział Gdański PTETiS
Referat nr 17

WIDZENIE KOMPUTEROWE OPARTE NA MNOGOŚCI WIDOKÓW

Adam Łukasz KACZMAREK[1]

1. Politechnika Gdańska; Wydział Elektroniki, Telekomunikacji i Informatyki
tel: 58 347 13 78 fax: 58 347 22 22 e-mail: adam.l.kaczmarek@eti.pg.gda.pl

Streszczenie: Artykuł poświęcony jest tematowi tworzenia map głębokości na podstawie obrazów z wielu kamer. Zwykle mapy głębokości oparte na widzeniu stereoskopowym wyznaczane są na podstawie obrazów z dwóch kamer. Artykuł przedstawia możliwości wykorzystania większej liczby kamer w celu zwiększenia dokładności map głębokości. Badania przedstawione w artykule ukierunkowane są na zastosowanie w autonomicznych robotach, będących w stanie samodzielnie przemieszczać się w środowisku miejskim. Do robotów takich zaliczane są między innymi samochody sterowane wyłącznie za pomocą systemów komputerowych. Artykuł przedstawia ponadto wielowarstwowe mapy głębokości, w których oprócz odległości obiektów pierwszoplanowych uwzględniana jest odległość obiektów położonych za nimi.

Słowa kluczowe: mapy głębokości, matryce kamer

1. WPROWADZENIE

Widzenie komputerowe polega na interpretowaniu przez systemy komputerowe obrazów z otaczającej nas rzeczywistości. W widzeniu komputerowym przeprowadzana jest analiza tego, co znajduje się na obrazach obiektów rzeczywistych przesyłanych do komputera. Wyznaczane są również mapy głębokości (ang. *depth maps*). Mapa głębokości określa odległość punktów znajdujących się na obrazie od miejsca położenia kamery lub aparatu, którym zostało wykonane zdjęcie.

Powszechnie stosowana metoda tworzenia map głębokości polega na wykonaniu dwóch zdjęć z kamer znajdujących się obok siebie [1]. Następnie porównuje się lokalizacje obiektów na tych dwóch obrazach. Z uwagi na fakt wykonania zdjęć z różnych miejsc, obiekty będą na obrazach znajdować się w różnych punktach oraz będą przesunięte względem siebie. W celu wykonania mapy głębokości tworzona jest najpierw mapa rozbieżności (ang. *disparity map*). Jeśli pewien obiekt zlokalizowany jest na pierwszym obrazie w punkcie x_1, y_1, a na drugim obrazie w punkcie x_2, y_1, to dla tego obiektu wartość rozbieżności jest równa $|x_1-x_2|$. Na podstawie rozbieżności, znając odległość, w jakiej umieszczone zostały kamery i ogniskową aparatu, można obliczyć mapę głębokości.

Zwiększenie dokładności map głębokości uzyskuje się zwykle przez zastosowanie bardziej rozbudowanych algorytmów dopasowania (ang. *matching*) lokalizacji obiektów znajdujących się na pierwszym obrazie, z ich lokalizacją na drugim obrazie. W tym artykule przedstawiony jest inny sposób, który polega on na zwiększeniu dokładności map głębokości przez zastosowanie większej liczby kamer. Dzięki temu uzyskuje się większą liczbę danych określających położenie obiektów. Wykorzystanie tych danych jest szczególnie istotne w zastosowaniach wymagających tworzenia map głębokości o wysokiej precyzji. Do takich zastosowań należy używanie tego rodzaju map na przykład w samochodach, które poruszają się po terenie zabudowanym bez kierowcy, lecz wyłącznie dzięki sterowaniu przez algorytmy komputerowe.

2. MAPY GŁĘBOKOŚCI DLA WIELU KAMER

Dotychczas prowadzone już były badania poświęcone temu tematowi. Tworzone były matryce kamer (ang. *camera array*) składające się z dużej liczby kamer umieszczanych w różnych konfiguracjach. Największa matryca kamer skonstruowana została na uniwersytecie Berkley [2]. Składała się ona ze 100 identycznych kamer. Stosowano różne konfiguracje, na przykład tworzono 8 rzędów po 12 kamer w każdym z nich. Matryca ta stosowana była głównie do detekcji szybko poruszających się obiektów.

Matryce kamer stosowane były również do tworzenia filmów trójwymiarowych [3]. W laboratoriach Mitsubishi opracowany został kompletny system rejestrowania obrazu trójwymiarowego oraz wyświetlania go za pomocą matrycy projektorów. Wykorzystanych zostało 16 kamer do rejestrowania obrazu.

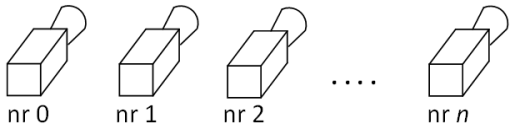
Do map głębokości używa się również mniejszej liczby kamer, w szczególności trzech [4]. Najczęściej stosowanym układem 3 kamer służących do obliczania mapy głębokości jest tzw. konfiguracja kąta prostego (ang. *right angle*). Trzecią kamerę umieszcza się pod jedną z kamer z pary kamer przeznaczonych do widzenia stereoskopowego. Williamson i Thorpe zastosowali układ trzech kamer przeznaczonych do sterowania autonomicznym samochodem kierowanym wyłącznie za pomocą algorytmów komputerowych [5]. Opracowali algorytm służący do wykrywania przeszkód znajdujących się na drodze. Algorytm ten był w stanie wykrywać przeszkody o rozmiarze większym niż 9 cm znajdujących się w odległości do 110 metrów.

Istnieje wiele metod służących do tworzenia map rozbieżności i głębokości [6]. Stosuje się ogólny podział tych metod na lokalne i globalne. Metody lokalne polegają na tym, że wartość odległości obliczana jest dla każdego punktu niezależnie od innych punktów. Natomiast w metodach

globalnych cała zawartość rysunku ma wpływ na wartość odległości w poszczególnych punktach mapy głębokości. Metody lokalne są zwykle szybsze i prostsze obliczeniowo niż metody globalne. Niniejszy artykuł poświęcony jest tworzeniu map głębokości za pomocą metod lokalnych.

Najbardziej rozpowszechniona lokalna metoda tworzenia map głębokości polega na tym, że dla danego punktu rysunku znajdowany jest punkt na drugim rysunku, który jest najbardziej podobny do punktu z pierwszego obrazu. Brany jest pod uwagę nie tylko pojedynczy punkt, lecz również jego otoczenie. Otoczenie to nazywane jest oknem (ang. *window*), w skład którego wchodzą punkty sąsiadujące z punktem badanym.

W celu zidentyfikowania najbardziej podobnych punktów stosuje się miary podobieństwa. Najczęściej stosowanymi są: miara SAD (*Sum of Absolute Differences*) oraz SSD (*Sum of Squared Differences*) [6]. W niniejszym artykule miary te są rozpatrywane w kontekście szeregu kamer położonych wzdłuż jednej linii i oddalonych od siebie w równych odległościach. Kamery numerowane są od 0 do *n*, gdzie *n* jest liczbą naturalną. Zatem liczba kamer wynosi *n*+1. Kamera numer 0 będzie kamerą referencyjną, względem której tworzone są mapy głębokości. Układ kamer przedstawiony jest na rysunku 1.



Rys. 1. Sposób numeracji szeregu kamer umieszczonych w równych odległościach

Równanie 1 przedstawia miarę SAD liczoną dla pary kamer, na którą składa się kamera referencyjna o numerze 0 i kamera oznaczona indeksem j.

$$SAD_j(x,y,d_j)=\sum_{k,l\in W}\left|I_0(x+k,y+l)-I_j(x-d_j+k,y+l)\right| \quad (1)$$

gdzie: *x,y* – współrzędne punktu, dla którego obliczana jest miara, I_0, I_j – wartość punktu o podanych współrzędnych na obrazach z kamer nr *0* oraz *j*, W - zakres okna, dla którego obliczana jest miara podobieństwa, d_j – wartość rozbieżności

Miara SAD opiera się na tym, że dla danego obszaru, dla którego miara ta jest liczona, sumowane są różnice wartości w poszczególnych punktach. Przy obliczaniu mapy głębokości stosuje się zwykle obrazy, w których punkty są określone w skali szarości, a nie obrazy kolorowe [6]. Zatem wartości w poszczególnych punktach określone w równaniu 1 funkcją *I*, odpowiadają natężeniu światła w danym punkcie obrazu. Kolejną miarę, jaką jest SSD, przedstawia równanie 2.

$$SSD_j(x,y,d_j)=\sum_{k,l\in W}\left(I_o(x+k,y+l)-I_j(x-d_j+k,y+l)\right)^2 \quad (2)$$

gdzie: oznaczenia takie jak we wzorze 1.

Miara SSD jest bardziej rozbudowana niż miara SAD. Obliczenie jej wymaga większej liczby operacji. Charakteryzuje się tym, że zamiast obliczania wartości bezwzględnej różnicy natężenia światła w punktach obrazów obliczana jest wartość tej różnicy podniesiona do kwadratu. Miara ta prowadzi do lepszych rezultatów, jednak kosztem większej liczby obliczeń.

W przypadku obydwu miar, poszukiwana jest wartość rozbieżności d_j, dla której miary te osiągają wartość najmniejszą. Wówczas przyjmuje się, że dla danego punktu z obrazu z kamery nr 0 został znaleziony odpowiadający mu punkt na obrazie *j*. Mapa rozbieżności w badanym punkcie przyjmuje wartość równą wartości d_j, dla której wartość miary była najmniejsza.

Miarę SSD uogólnia się na wiele kamer. Realizowane jest to przez zdefiniowanie funkcji SSSD (*Sum of SSD*). Polega ona na zsumowaniu wartości SSD dla wszystkich par kamer od pary 0, 1 po parę 0, *n*. W wyniku tego powstaje miara, w której uwzględniane są obrazy ze wszystkich kamer. Równanie 3 przedstawia funkcję SSSD.

$$SSSD(x,y,d)=\sum_{j=1}^{n}SSD(x,y,d_j) \quad (3)$$

Miara ta posiada jednak wadę polegającą na tym, że w przypadku, gdy na obrazie są powtarzające się wzorce, to osiąga wiele minimów, a nie jedynie minimum dla właściwej wartości rozbieżności *d*. W związku z tym wprowadzona została inna funkcja nazwana SSSD-in-inverse-distance [7]. Jest ona określona równaniem 4.

$$SSSD(x,y,\zeta)=\sum_{j=1}^{n}\sum_{k,l\in W}\left(I_o(x+k,y+l)-I_j(x+B_jF\zeta+k,y+l)\right)^2 \quad (4)$$

gdzie: ζ – odwrotność odległości między kamerą a punktem przedstawionym na obrazie, B_j – odległość między kamerą 0 a kamerą *j*, *F* – wartość ogniskowej kamer

Funkcja przedstawiona równaniem 4 charakteryzuje się tym, że jej trzecim argumentem nie jest wartość rozbieżności, lecz odwrotność odległości między kamerą, a punktem przedstawionym na obrazie. Dzięki temu zredukowany jest problem występowania wielu minimów występujący w funkcji przedstawionej równaniem 3.

3. JAKOŚĆ MAP GŁĘBOKOŚCI

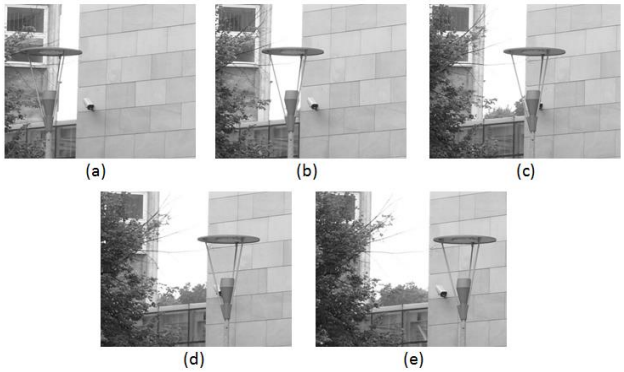
Ocena jakości mapy głębokości przeprowadzana jest na podstawie porównania jej z wzorcową mapą głębokości (ang. *ground truth)* odzwierciedlającą rzeczywiste odległości między kamerą a obiektami przedstawionymi na obrazie. Stosowane są dwie metody porównywania mapy badanej z mapą wzorcową [1].

W pierwszej z metod określana jest pewna dopuszczalna wartość błędu dla punktów składających się na mapę głębokości. Jeśli różnica między wartością w pewnym punkcie mapy głębokości i wartością w tym samym punkcie mapy wzorcowej jest mniejsza od zadanego progu, to przyjmuje się, że dla danego punktu odległość została obliczona prawidłowo. W przeciwnym wypadku wartość punktu mapy głębokości jest określana jako obliczona błędnie. Następnie obliczany jest odsetek punktów mapy głębokości obliczonych poprawnie. Odsetek ten określa poziom jakości mapy głębokości.

Druga metoda szacowania jakości mapy głębokości opiera się na średniej kwadratowej. Dla każdego punktu obliczana jest różnica między wartością mapy głębokości i mapy wzorcowej. Następnie wyznaczana jest średnia kwadratowa wszystkich obliczonych różnic.

4. WIELOWARTSWOWE MAPY GŁĘBOKOŚCI

Rysunek 2 przedstawia sekwencję zdjęć wykonanych przez autora niniejszego artykułu przedstawiających fragmenty dwóch budynków Wydziału Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej. Zdjęcia zostały wykonane z 5 różnych ujęć. Miejsca, z których wykonane zostały te zdjęcia, były od siebie równo oddalone. Tego rodzaju sekwencja zdjęć jest punktem wyjścia do badań prowadzonych przez autora.



Rys. 2. Sekwencja zdjęć budynków Wydziału Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej

Przy tworzeniu map głębokości za pomocą funkcji *SSSD-in-inverse-distance* przedstawionej równaniem 4 bierze się pod uwagę wszystkie obrazy jednocześnie. Możliwe jest również zinterpretowanie sekwencji zdjęć jako zbioru par zdjęć z różnych kamer. Jeśli użytych zostało *n+1* kamer, to istnieje *n(n+1)/2* par zdjęć. Jeśli dąży się do tego, żeby wyznaczyć mapę głębokości dla obiektów znajdujących się na obrazie z kamery 0, to nie wykorzystuje się wszystkich par kamer. Używa się jedynie par, na które składa się kamera nr 0 oraz jedna z innych kamer. Dla każdego punktu przedstawionego na obrazie z kamery 0 określana jest odległość od kamery na podstawie pozostałych zdjęć.

Jednak na podstawie sekwencji zdjęć, takiej jak przedstawiona na rysunku 2, możliwe jest nie tylko stworzenie mapy odległości, ale również wyodrębnienie obiektów znajdujących się na pierwszym planie od obiektów położonych dalej. Człowiek patrząc na określone zdjęcie, na którym obiekt na pierwszym planie przysłania obiekty położone dalej, intuicyjnie przypuszcza, jakie obiekty znajdują się za obiektem przysłaniającym te, które są bardziej odległe. Algorytmy komputerowe do pewnego stopnia również są w stanie interpretować zawartość obrazów i oszacować, jakie obiekty znajdują się za obiektem pierwszoplanowym. Operacja taka jest jednak złożona obliczeniowo. Mniej wymagająca jest operacja usunięcia z pierwszego planu zdjęcia obiektu, który przysłania inny obiekt, na postawie sekwencji zdjęć takiej, jak przedstawiona na rysunku 2.

Połączenie obrazów pozwalające na pominięcie obiektów pierwszoplanowych zrealizowane może być w postaci zaproponowanej w tym artykule wielowarstwowej mapy głębokości. Jest ona tworzona w następujących krokach:

- Utworzenie map głębokości dla wszystkich obrazów
- Podział mapy głębokości obrazu z kamery nr 0 na warstwy
- Wyznaczenie obszarów widocznych z innej kamery niż kamera 0
- Dołączenie map głębokości obszarów uzyskanych w punkcie 3 do warstw uzyskanych w punkcie 2

4.1. Utworzenie map głębokości dla wszystkich obrazów

W pierwszym kroku tworzenia wielowarstwowej mapy głębokości obliczane są mapy dla wszystkich obrazów. Mogą one być obliczane jedynie na podstawie pary obrazów. Możliwe jest jednak również uwzględnienie sekwencji obrazów. Na przykład mapa głębokości dla obrazu z kamery nr 2 może być tworzona na podstawie pary, w skład której wchodzi obraz z kamery nr 2 i dowolny inny obraz. Do tworzenia mapy głębokości zastosowany może zostać algorytm odpowiedni dla danego rodzaju obrazów. Najczęściej stosowane są algorytmy oparte na miarach SAD i SSD przedstawionych w rozdziale 2.

Do tworzenia map głębokości dla wszystkich obrazów można również zastosować algorytm *SSSD-in-inverse-distance*. Na przykład dla obrazu z kamery 2 można utworzyć mapę głębokości na podstawie obrazów z kamer od 2 do *n*. Możliwe są również inne konfiguracje. Wybór metody obliczenia map głębokości zależny jest od rodzaju obrazów oraz oczekiwanego stopnia dokładności map.

4.2. Podział mapy głębokości obrazu z kamery nr 0 na warstwy

Po uzyskaniu w poprzednim kroku map głębokości dla wszystkich obrazów, w kolejnym kroku przetwarzana jest mapa obrazu z kamery nr 0. Jest ona dzielona względem określonych na tym obrazie odległości lub zakresów odległości.

Wszystkie punkty mapy o tej samej wartości składają się na jedną warstwę. W przypadku, gdy jest zbyt wiele różnych dopuszczalnych wartości punktów, możliwe jest połączenie warstw, przez co pojedyncza warstwa nie będzie zawierać tylko punktów o określonej wartości, lecz punkty z pewnego zakresu wartości. Podział mapy głębokości na warstwy służy zidentyfikowaniu obszarów obrazu znajdujących się na pierwszym planie oraz obszarów drugoplanowych.

4.3. Wyznaczenie obszarów widocznych z innej kamery niż kamera 0

Uzyskane w poprzednim kroku warstwy obrazu z kamery 0, które odpowiadają obiektom pierwszoplanowym, wyznaczają obszary na pozostałych obrazach, które pokazują obiekty położone za obiektami pierwszoplanowymi obrazu z kamery nr 0. Dla obiektów pierwszoplanowych wartość odległości jest na mapie głębokości mniejsza niż obiektów położonych dalej. Jednocześnie jeśli operowalibyśmy mapą rozbieżności, to rozbieżność będzie większa dla obiektów pierwszoplanowych. Za obiektami pierwszoplanowymi położone są obiekty o mniejszej wartości rozbieżności i większej wartości odległości.

Na podstawie lokalizacji obiektów pierwszoplanowych na obrazie z kamery nr 0 można określić obszary na pozostałych obrazach przedstawiające obiekty położone za obiektem pierwszoplanowym. Jest to możliwe do określania, jeśli znane są odległość kamer od siebie, wartość ogniskowej kamery oraz odległość obiektów pierwszoplanowych od kamery.

Obszar za obiektem pierwszoplanowym może nie być w całości widoczny na innych kamerach. Obszar ten będzie

najbardziej widoczny na kamerach najbardziej oddalonych od kamery nr 0, o ile inny obiekt nie przesłoni tego obszaru. Po wyznaczeniu tych obszarów wykonywany jest kolejny krok mający na celu stworzenie wielowarstwowej mapy głębokości.

4.4. Połączenie map głębokości

W ostatnim kroku tworzenia wielowarstwowej mapy głębokości łączone są dane uzyskane w kroku 2 i w kroku 3. Następuje rozszerzenie warstw. Do warstw mapy głębokości obrazu z kamery nr 0 dodawane są fragmenty map głębokości pozostałych obrazów odpowiadające obszarom wyznaczonym w kroku 3, czyli tych obszarów, które obrazują obiekty położone za obiektem pierwszoplanowym na obrazie z kamery nr 0.

Dodawanie tych fragmentów map głębokości odbywa się w taki sposób, że do każdej warstwy uzyskanej w kroku 2 dodawane są obszary odpowiadające tej samej odległości od kamery, co dana warstwa. W ten sposób uzyskiwane są warstwy, w których każda odpowiada jednej odległości od kamery. Warstwy te, z naniesionymi danymi uzyskanymi w kroku 3, tworzą wielowarstwową mapę głębokości.

4.5. Ocena jakości wielowarstwowych map głębokości

Jakość wielowarstwowych map głębokości może być oceniana w podobny sposób, jak jakość typowych map głębokości opisana w rozdziale 3 niniejszego artykułu. Istnieje jednak istotna różnica w rodzaju wzorca, jaki powinien być używany do takiej oceny. W przypadku typowych map głębokości wzorzec zawiera odległość od kamery wszystkich punktów znajdujących się na obrazie z kamery nr 0. Wielowarstwowa mapa głębokości zawiera więcej punktów z racji tego, że zawiera również informacje o odległości obiektów niewidocznych. Przy ocenie jakości wielowarstwowych map głębokości konieczne jest określenie wielowarstwowego wzorca uwzględniającego odległości tych obszarów.

Wzorcowa wielowarstwowa mapa głębokości nie może być skonstruowana tak, że na każdej warstwie będą wszystkie punkty odpowiadające rzeczywistym odległościom. Nie jest to możliwe, ponieważ część obszarów niewidocznych z kamery nr 0 nie będzie również widoczna z pozostałych kamer. Wzorcowa wielowarstwowa mapa odległości powinna uwzględniać jedynie te obszary, dla których powinny zostać wyznaczone odległości na podstawie obrazów ze wszystkich kamer.

5. PODSUMOWANIE

Zastosowanie wielowarstwowych map głębokości jest szerokie. Autonomiczne roboty, w szczególności pojazdy sterowane algorytmami komputerowymi, na podstawie wielowarstwowej mapy odległości mogą określać nie tylko położenie obiektów znajdujących się bezpośrednio przed nimi lecz również obiekty położone dalej. Dzięki temu, na przykład podczas manewru omijania pewnego obiektu, można również uwzględnić manewr omijania równocześnie kolejnego obiektu, położonego w większej odległości niż pierwszy omijany obiekt.

6. BIBLIOGRAFIA

1. Scharstein D., Szeliski R.: A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, International Journal of Computer Vision, Kluwer Academic Publishers April 2002, Volume 47, Issue 1-3, s. 7-42, ISSN: 0920-5691.
2. Wilburn B., Joshi N., Vaish V., Talvala E., Antunez E., Barth A., Adams A., Horowitz M., Levoy M.: High performance imaging using large camera arrays, ACM Transactions on Graphics, Volume 24, No 3, ACM July 2005, s. 765-776, ISSN 0730-0301.
3. Matusik W., Pfister H.: 3D TV A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes, ACM SIGGRAPH, August 2004, Volume 23, Issue 3, s. 814-824, ISSN: 0730-0301.
4. Agrawal M., Davis L. S.: Trinocular Stereo Using Shortest Paths and the Ordering Constraint, International Journal of Computer Vision, Volume 47, Issue 1-3, Kluwer Academic Publishers April 2002, s. 43-50, ISSN: 0920-5691.
5. Williamson T., Thorpe C.: A trinocular stereo system for highway obstacle detection, Proceedings of IEEE International Conference on Robotics and Automation, USA Detroit 1999, ISBN: 0-7803-5180-0.
6. Lazaros N., Sirakoulis G. C., Gasteratos A.: Review of Stereo Vision Algorithms: From Software to Hardware, International Journal of Optomechatronics, Volume 2, Taylor & Francis Group LLC 2008, s. 435–462, ISSN: 1559-9612.
7. Okutomi M., Kanade T.: A Multiple-Baseline Stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 15, No. 4, IEEE Computer Society April 1993, s. 353-363, ISSN: 0162-8828.

COMPUTER VISION ON THE BASIS OF MULTIPLE VIEWS

Key-words: depth map, camera array

The paper is concerned with creating depth maps on the basis of images from a set of multiple cameras. Most often depth maps are prepared on the basis of only two cameras. The paper presents possibilities of improving depth maps by taking advantage of a greater number of cameras. The research presented in this paper is intended for use in autonomous robots able to navigate in cities without human control. These kind of robots includes cars controlled by computer algorithms. Moreover the paper presents multilayer depth maps. These maps include both the distance of objects located in the foreground and objects located behind them.