

ZASTOSOWANIE ALGORYTMU MSA (MULTIPLE SIMILAR AREAS) DO WYZNACZANIA MAP GŁĘBI W WIELOWIDOKOWYCH SYSTEMACH WIDZENIA KOMPUTEROWEGO

Adam L. KACZMAREK

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

Streszczenie: Artykuł podejmuje temat pozyskiwania map głębi (ang. depth map) na podstawie zdjęć z wielu kamer w wyniku widzenia stereoskopowego. Mapa głębi zawierająca odległości od obiektów będących w zasięgu widzenia kamer pozyskana może zostać na podstawie zdjęć z co najmniej dwóch kamer pełniących funkcję kamery stereoskopowej. W mapach głębi pozyskanych w ten sposób występują jednak błędy. Artykuł dotyczy metod redukcji błędów dzięki zwiększeniu liczby kamer. W artykule przedstawione zostało zastosowanie algorytmu MSA (Multiple Similar Areas) przeznaczonego do układu kamer o nazwie EBMCs (Equal Baseline Multiple Camera Set). Układ taki składa się z kamery środkowej oraz wielu kamer bocznych. Wykonywanie zdjęć trójwymiarowych za pomocą układu wielu kamer jest alternatywą do stosowania innego rodzaju urządzeń takich jak skanery światła strukturalnego oraz skanery laserowe typu LIDAR.

Słowa kluczowe: mapa głębi, widzenie stereoskopowe, systemy wielokamerowe, łańcuch kamer.

1. WPROWADZENIE

Dotychczas opracowanych zostało wiele rodzajów urządzeń pozwalających na tworzenie zdjęć trójwymiarowych. Do urządzeń takich należą skanery światła strukturalnego [1], kamery czasu przelotu TOF (time-of-flight cameras) [2], urządzenia typu LIDAR (Light Detection and Ranging) [3], kamery stereoskopowe [4] oraz układy wielu kamer [4]. Hartley i Zisserman opisali w szerokim zakresie metody otrzymywania map głębi na podstawie zdjęć z układu wielu kamer [5].

Działanie skanerów światła strukturalnego polega na emitowaniu wzorca świetlnego o ściśle określonym kształcie, a następnie przeprowadzeniu analizy załamania światła na badanym obiekcie. Często stosowanym wzorcem świetlnym są podłużne, równoległe do siebie prążki. Powszechnie stosowanym skanerem światła strukturalnego jest urządzenie typu MS Kinect operujące w zakresie światła podczerwonego. Innymi rodzajami urządzeń są kamery czasu przelotu TOF oraz skanery typu LIDAR. Urządzenia te, podobnie jak skanery światła strukturalnego, emitują wiązkę świetlną, jednak jest to wiązka laserowa. Kamery TOF i LIDAR badają odbicie promienia laserowego od powierzchni badanego obiektu i na tej podstawie określają odległość między obiektem a urządzeniem pomiarowym. Pomiar wiązką lasera w określonym punkcie obiektu jest bardzo dokładny. Urządzenia tego typu posiadają jednak wadę polegającą na tym, że pozyskują kształt obiektu w niskiej rozdzielczości. Wynika to z faktu, że pomiar

laserowy odbywa się jedynie w wielu pojedynczych punktach badanego obiektu, nie zaś na całej jego powierzchni.

Zarówno skanery światła strukturalnego, kamery czasu przelotu oraz skanery LIDAR emitują wiązkę światła w celu wykonania pomiaru. W związku z tym wpływ na pomiar za pomocą tych urządzeń ma oświetlenie pochodzące z innych źródeł światła. Nie stanowi to problemu, gdy można kontrolować warunki oświetlenia, na przykład w pomieszczeniu zamkniętym. Jednak pomiar jest zaburzony, jeśli przeprowadzany jest w intensywnym świetle naturalnym. Dotyczy to przede wszystkim skanerów światła strukturalnego i kamer czasu przelotu. Dodatkowo szczególnie problematyczne jest wykonanie za pomocą tych urządzeń zdjęcia trójwymiarowego obiektów o dużych rozmiarach, takich jak np. budynki. Wynik pomiaru jest albo niskiej rozdzielczości: w przypadku kamer TOF i skanerów LIDAR, albo w przypadku skanerów światła strukturalnego istnieje konieczność wygenerowania bardzo silnej wiązki światła, która jest w stanie oświetlić budynek.

Problemy te nie występują, jeśli obraz trójwymiarowy jest generowany za pomocą kamer stereoskopowych oraz układów wielokamerowych. Tego rodzaju układy kamer wykorzystywane są do wykonywania zdjęć obiektów z różnych punktów widzenia. Lokalizacja obiektów i ich części na zdjęciach będzie różna na w zależności od tego, którą kamerą zostało wykonane zdjęcie. Im obiekt będzie znajdował się bliżej kamer, tym większa będzie względna różnica w lokalizacji na zdjęciach z różnych kamer tego samego obiektu. Na podstawie różnicy w położeniu obiektów na zdjęciach możliwe jest obliczenie odległości między kamerami a obiektem. Jest to możliwe, jeśli dodatkowo znane są parametry kamery stereoskopowej takie, jak wartości ogniskowych soczewek oraz odległość między kamerami.

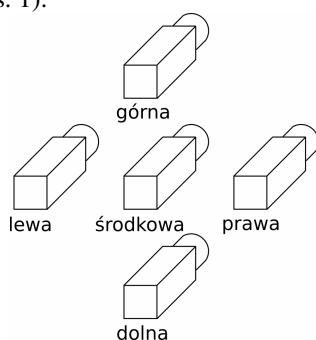
Prace badawcze przedstawione w tym artykule dotyczą metod poprawiania dokładności zdjęć trójwymiarowych poprzez zwiększenie liczby kamer użytych do wykonania zdjęć. Wykonanie zdjęcia trójwymiarowego jest możliwe, jeśli dysponuje się jedynie parą kamer. Jednak dokładność zdjęcia można zwiększyć, korzystając z dodatkowych kamer wykonujących zdjęcia z innych punktów widzenia [6]. W dotychczasowych badaniach prowadzonych przez autora tego artykułu analizowane było zastosowanie układu pięciu kamer w ściśle określonym położeniu służącego do wykonywania zdjęć trójwymiarowych. Do tego celu opracowany został przez autora niniejszego artykułu algorytm o nazwie MSA (Multiple Similar Areas). Opisany on został w publikacji [7]. Algorytm dostosowany został

celowo do charakterystyki badanego układu pięciu kamer. W tym artykule przedstawiona jest propozycja uogólnienia algorytmu MSA na inne układy kamer.

Do oryginalnych osiągnięć przedstawionych w artykule należy: propozycja zastosowania algorytmu MSA do układu kamer zawierającego kamerę centralną oraz dowolną liczbę kamer bocznych, przedstawienie modyfikacji algorytmu MSA pozwalających na użycie algorytmu z tego rodzaju układem kamer oraz przedstawienie problemów, jakie wystąpiłyby w przypadku zastosowania algorytmu do układu zawierającego dowolnie dużą liczbę kamer umieszczonych w różnych odległościach od kamery centralnej.

2. ALGORYTM MSA (MULTIPLE SIMILAR AREAS)

Algorytm MSA będący przedmiotem tego artykułu przeznaczony jest do stosowania z układem pięciu kamer w konfiguracji EBMCS (Equal Baseline Multiple Camera Set) [6][7][8]. Układ ten składa się z kamery środkowej oraz czterech kamer bocznych umieszczonych dookoła kamery środkowej (Rys. 1).



Rys. 1. Układ kamer w konfiguracji EBMCS (Equal Baseline Multiple Camera Set)

Zbiór kamer w konfiguracji EBMCS jest interpretowany w algorytmie MSA jako zbiór czterech kamer stereoskopowych. Każda z nich składa się z kamery środkowej oraz jednej z kamer bocznych. W każdej parze kamera środkowa pełni funkcję kamery referencyjnej. Jest to kamera, względem której obliczana jest odległość do obiektów będących w polu widzenia kamer. Konfiguracja EBMCS charakteryzuje się tym, że odległości między kamerą środkową a kamerą boczną są takie same dla wszystkich rozpatrywanych par kamer

Algorytm MSA przeznaczony jest do tworzenia map głębi. Są to obrazy, w których wartości punktów odpowiadają odległości między urządzeniem pomiarowym a badanymi obiektami. W celu uzyskania map głębi oblicza się mapy rozbieżności odzwierciedlające różnice w położeniu obiektów na zdjęciach bocznych względem ich położenia na zdjęciu środkowym [7].

Podstawą algorytmu MSA jest to, że znajduje on ciągi podobnych do siebie punktów występujących jednocześnie na obrazach ze wszystkich kamer. Dla każdego punktu obrazu referencyjnego algorytm wyszukuje odpowiadające mu punkty na obrazach bocznych w pewnych obszarach obrazu, w których istnieje taka możliwość, że punkty te będą występować. Położenie obiektu na zdjęciach z kamer bocznych jest przesunięte względem położenia tego obiektu na zdjęciu z kamery środkowej w zależności od lokalizacji kamery bocznej. W każdej z rozpatrywanych par kamer występuje przesunięcie w inną stronę. W algorytmie MSA obliczane są funkcje m dane wzorami (1) [7]:

$$\begin{aligned} m_1(x, y, d) &= S_1(x, y, x - d, y) \\ m_2(x, y, d) &= S_2(x, y, x, y + d) \\ m_3(x, y, d) &= S_3(x, y, x + d, y) \\ m_4(x, y, d) &= S_4(x, y, x, y - d) \end{aligned} \quad (1)$$

gdzie: x, y – współrzędne punktu, d – wartość rozbieżności, S – miara podobieństwa dwóch punktów.

Funkcja S używana we wzorach (1) określa, czy dwa punkty są do siebie podobne pod względem koloru lub natężenia światła. Funkcja m określa parametry funkcji S w zależności od rozpatrywanej pary kamer. Indeksy 1, 2, 3 i 4 funkcji odpowiadają parom kamer zawierających kamerę odpowiednio: prawą, górną, lewą oraz dolną. W funkcji m pierwszym i drugim argumentem (x, y) jest położenie punktu, dla którego obliczana jest głębia, natomiast ostatni argument jest wartością rozbieżności (d) będącą różnicą między położeniem obiektu na obrazku środkowym a jego położeniem na obrazku bocznym. W funkcji S dwa pierwsze argumenty są takie same, jak w funkcji m , natomiast trzeci i czwarty argument określa lokalizację badanego punktu na obrazku bocznym. W przypadku zastosowania algorytmu do układów kamer innych niż konfiguracja EBMCS istnieje konieczność modyfikacji wzorów (1). Sposób ich modyfikacji został przedstawiony w rozdziale następnym.

Algorytm MSA wylicza wartości funkcji m w całym zakresie badanych wartości rozbieżności d , a następnie znajduje najdłuższe ciągi kolejnych wartości rozbieżności, dla których spełnione są powyższe równania. Szczegóły algorytmu opisane są w publikacji [7]. Algorytm użyty dla pięciu kamer prowadzi do wyników zawierających 63% procent mniej błędów niż wyniki uzyskane dla dwóch kamer [7]. Dotychczas powstało wiele algorytmów przeznaczonych do stereoskopii na podstawie zdjęć z pary kamer, jednak algorytmy przeznaczone dla układów wielu kamer rozwijane były do tej pory w niewielkim zakresie. Algorytm MSA umożliwia zwiększenie jakości oszacowania głębi dzięki zwiększeniu liczby kamer użytych do pomiaru.

3. ZASTOSOWANIE ALGORYTMU MSA

Algorytm MSA przeznaczony jest do tworzenia map głębi na podstawie zdjęć z układu pięciu kamer w konfiguracji EBMCS. Algorytm ten może być jednak dostosowany do układów zawierających inną liczbę kamer, które dodatkowo rozmieszczone są w inny sposób niż w konfiguracji EBMCS. Konieczne jest jednak wprowadzenie modyfikacji do algorytmu. Modyfikacje te polegają przede wszystkim na uogólnieniu kroków algorytmu, które dostosowane są wyłącznie do charakterystyki EBMCS.

Zastosowanie algorytmu MSA do układów kamer w konfiguracjach innych niż EBMCS przedstawione zostało w dwóch wariantach. Pierwszy wariant dotyczy układu kamer zawierających kamerę centralną oraz dowolną liczbę kamer bocznych oddalonych w równej odległości od kamery centralnej. Wariant drugi opisuje układ kamer, w którym występuje centralna kamera referencyjna oraz kamery boczne oddalone od kamery centralnej w różnej odległości.

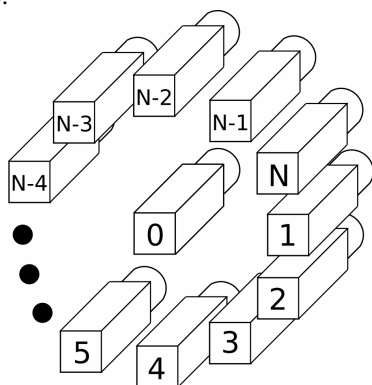
3.1. Układ z kamerą centralną i równo oddalonymi kamerami bocznymi

Algorytm MSA stosowany był dotychczas jedynie z układem kamer w konfiguracji EBMCS, zawierającym pięć urządzeń. Jednak liczbę urządzeń w tej konfiguracji można

zwiększyć, zachowując jednocześnie jej właściwości. Układ zawierający dowolnie dużą liczbę kamer jest zgodny z warunkami układu w konfiguracji EBMCS, jeśli spełnia następujące warunki:

- występuje jedna kamera centralna oraz kamery boczne,
- układ rozpatrywany jest jako zbiór kamer stereoskopowych składających się z kamery centralnej oraz jednej z kamer bocznych,
- w każdym rozpatrywanej parze kamer kamera centralna pełni funkcję kamery referencyjnej,
- odległości między każdą kamerą boczną a kamerą centralną są sobie równe.

Struktura tego rodzaju układu kamer przedstawiona jest na rysunku 2:



Rys. 2. Konfiguracja EBMCS zawierająca dowolnie dużą liczbę kamer

W celu uogólnienia algorytmu MSA na konfigurację EBMCS z dowolnie dużą liczbą kamer konieczne jest przeprowadzenie modyfikacji kroku algorytmu, w którym dla wszystkich czterech rozpatrywanych par kamer sprawdzane są odpowiadające sobie fragmenty zdjęć z różnych kamer. Krok ten przedstawiony jest w postaci wzorów (1) podanych w poprzednim rozdziale. W celu uogólnienia funkcji na dowolną liczbę kamer w konfiguracji EBMCS wystarczy zamiast wzorów (1) zastosować wzór (2):

$$m_n(x, y, d) = S_n(x, y, x - d \cdot \cos(\alpha), y + d \cdot \sin(\alpha)) \quad (2)$$

gdzie: α – kąt pomiędzy odcinkiem łączącym kamerę środkową z kamerą boczną a półprostą rozpoczynającą się w miejscu kamery środkowej i skierowaną w prawo, n – indeks kamery bocznej, pozostałe oznaczenia takie, jak we wzorach (1).

Mimo podobieństwa wzoru (2) do wzorów (1) istnieją między zastosowaniem tych wzorów pewne różnice. W przypadku wzorów przeznaczonych dla EBMCS z pięcioma kamerami funkcje S wskazują punkty obrazka o współrzędnych mających wartości naturalne. Należy zwrócić uwagę, że obrazek przetwarzany w formie cyfrowej nie zawiera danych ciągłych, lecz jest zbiorem punktów. Gdy stosowany jest wzór (2), wówczas argumenty $x - d \cos(\alpha)$ i $x + d \cos(\alpha)$ nie przyjmują wartości liczb naturalnych, jeśli stosowane są dla pary kamer, w której kamera boczna jest umieszczona pod innym kątem niż 0° , 90° , 180° lub 270° . Współrzędne, które nie mają wartości naturalnych, wskazują pewien obszar pomiędzy sąsiadującymi ze sobą punktami obrazka.

Powoduje to konieczność określenia wartości punktu w tym właśnie obszarze. Można tego dokonać na dwa następujące sposoby:

- a) poprzez interpolację wartości na podstawie wartości w punktach sąsiadujących,
- b) poprzez przyjęcie, że wartość w określonych współrzędnych jest równa wartości punktu położonego najbliższej tych współrzędnych.

W przypadku stosowania algorytmu MSA możliwe jest zastosowanie obydwu tych metod określania wartości punktów w współrzędnych nienaturalnych. Jednak metoda druga, w której przyjmowana jest wartość punktu znajdującego się najbliższej punktu o danych współrzędnych wymaga mniej obliczeń, stąd jest ona preferowana. Dodatkowo określanie wartości punktu na podstawie wszystkich punktów sąsiadujących, na przykład poprzez wyliczenie wartości średniej, wiązałoby się z efektem rozmycia obrazu, co jest niewskazane.

Opisana modyfikacja wzorów jest najbardziej znaczącą zmianą w algorytmie MSA konieczną do zastosowania algorytmu do układu kamer EBMCS z dowolną liczbą kamer. Zmiany konieczne do wprowadzenia w innych krokach algorytmu polegają jedynie na tym, że algorytm zamiast korzystać z czterech par kamer wykorzystuje ich większą liczbę.

3.2. Układ z kamerą centralną i dowolnie oddalonymi kamerami bocznymi

Drugim rodzajem układu kamer, do którego dostosować można algorytm MSA jest konfiguracja, w której występuje kamera środkowa i kamery boczne, jednak odległość między kamerami bocznymi a środkową jest różna. Układ taki, podobnie jak w przypadku układu opisanego w rozdziale 3.1., również jest traktowany jako zbiór kamer stereoskopowych, w skład których zawsze wchodzi kamera środkowa. Układ tego rodzaju nie może być nazywany układem typu EBMCS, ponieważ nie jest spełniony warunek równych odległości między parami kamer branymi pod uwagę podczas wyznaczania mapy głębi. Mimo tego z układem może zostać zastosowany algorytm MSA, w którym wprowadzone zostały modyfikacje.

Problem układów kamer, w których występują różne odległości w poszczególnych parach kamer, był analizowany w kontekście użycia łańcuchów kamer (ang. *camera array*) do wyznaczania map głębi [4][9]. W łańcuchu takim występuje kamera początkowa oznaczana numerem 0 oraz kolejne kamery. W algorytmach przeznaczonych do widzenia stereoskopowego łańcuch kamer przetwarzany jest tak, jakby stanowił zbiór par kamer składających się z kamery 0 oraz z innej kamery z łańcucha. Przy wyznaczaniu głębi za pomocą takiego zbioru występuje problem polegający na tym, że na zdjęciach z różnych par kamer występują różne wartości rozbieżności dla obiektów oddalonych w takiej samej odległości od łańcucha. Jeśli dany jest pewien obiekt widoczny z pary kamer, to jeśli zwiększy się odległość między kamerami, to na zdjęciach uzyskanych z tej pary wystąpią większe rozbieżności niż wtedy, gdy kamery były umieszczone w mniejszej odległości. Różne rozbieżności na różnych parach kamer występujące dla tych samych obiektów utrudniają scalanie danych pozyskanych z różnych par kamer.

Okutomi i Kanade zaproponowali rozwiązanie tego problemu opierające się na koncepcji, że zamiast wyznaczać głębię względem wartości rozbieżności była ona

wyznaczania względem odwrotności odległości do badanych obiektów. Niezależnie od użytej pary kamer z łańcucha kamer odległość obiektów od pary będzie identyczna, ponieważ jest ona liczona względem odległości od kamery nr 0 wchodzącej w skład każdej analizowanej pary kamer. Rozwiązanie zaproponowane przez Okutomi i Kanade opiera się na spostrzeżeniu, że odwrotność odległości między kamerą nr 0 a obiektem fotografowanym jest dla każdej pary kamer wprost proporcjonalna do wartości rozbieżności uzyskanej za pomocą tej pary. W celu wyznaczenia rozbieżności konieczne jest jednak wykorzystanie danych dotyczących ogniskowych soczewek kamer oraz odległości między kamerami. Wartości te są jednak niezmiennie w danym układzie kamer. Zależność między odwrotnością odległości a rozbieżnością podana jest wzorem (3) [9]:

$$\zeta = \frac{d}{B \cdot F} \quad (3)$$

gdzie: ζ – odwrotność odległości, d – rozbieżność, B – odległość między kamerami, F – ogniskowa.

Stosowanie odwrotności odległości rozwiązuje problem różnych odległości między kamerami, jednak powoduje wydłużenie i skomplikowanie obliczeń, dlatego jest to metoda stosowana jedynie do łańcuchów kamer. Metoda ta jest jednak również adekwatna do zastosowania w przypadku użycia algorytmu MSA do układu wielu kamer oddalonych od siebie w różnej odległości. Funkcja m dana równaniem (2) zamiast argumentu d będącego rozbieżnością obliczana będzie względem argumentu ζ odpowiadającego odwrotności odległości. Stosownie do wartości tego argumentu określane będzie za pomocą funkcji S , czy w pewnym obszarze obrazków uzyskanych ze zdjęć bocznych występują fragmenty podobne do fragmentów z obrazka środkowego.

4. PODSUMOWANIE

W dalszych pracach autora poświęconych układom wielokamerowym planowane jest przeprowadzenie eksperymentów z algorytmem MSA zastosowanym do układów kamer zaproponowanych w tym artykule. Prowadzenie prac w tym zakresie jest szczególnie istotne ze względu na szeroki potencjał zastosowania tej technologii. W systemach komputerowych układy wielu kamer były dotychczas stosowane w ograniczonym zakresie. Jednak szerokie ich wykorzystanie możliwe jest dzięki rozwojowi prac badawczych poświęconych autonomicznym pojazdom

zdolnym do przemieszczania się bez ingerencji człowieka. W pojazdach takich istnieje konieczność zidentyfikowania przeszkód znajdujących się dookoła oraz odległości od tych przeszkód. Realizowane jest to między innymi dzięki użyciu kamer. Celem układu wielu kamer zaproponowanym w tym artykule jest zwiększenie dokładności pomiaru odległości między kamerami, a obiektami znajdującymi się w zasięgu ich widzenia poprzez zwiększenie liczby urządzeń badawczych. Celem dalszych prac autora będzie wyznaczenie liczby kamer, przy której dodanie kolejnego urządzenia nie powoduje zwiększenia dokładności pomiaru.

5. BIBLIOGRAFIA

1. Gupta M., Yin Q., Nayar S. K.: Structured light in sunlight. In 2013 IEEE International Conference on Computer Vision, Sydney 2013, s. 545-552.
2. Kazmi W., Foix S., Alenya G.: Plant leaf imaging using time of flight camera under sunlight, shadow and room conditions, IEEE Int. Symp. on Robotic and Sensors Environments (ROSE), Magdeburg Germany 2012, s. 192-197.
3. Tse R. O., Gold C., Kidner D.: 3D City Modelling from LIDAR Data, Advances in 3D Geoinformation Systems, Springer Berlin Heidelberg, Springer-Verlag Berlin Heidelberg 2008, s. 161-175.
4. Kaczmarek A. L.: Widzenie komputerowe oparte na mnogości widoków, Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej 36, Gdańsk 2013, s.85-88.
5. Hartley R. I., Zisserman A.: Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.
6. Park J.-I., Inoue S.: Acquisition of sharp depth map from multiple cameras, Signal Processing: Image Communication, 14(12), Elsevier 1998, s.7-19.
7. Kaczmarek A. L.: Improving depth maps of plants by using a set of five cameras, Journal of Electronic Imaging 24(2), SPIE and IS&T 2015, s. 023018.
8. Kaczmarek A. L.: Influence of Aggregating Window Size on Disparity Maps Obtained from Equal Baseline Multiple Camera Set (EBMCS), Image Processing and Communications Challenges 8: 8th International Conference, IP&C 2016 Bydgoszcz, Poland, September 2016 Proceedings, Springer International Publishing, s. 187-194, 2017.
9. Okutomi M., Kanade T.: A multiple-baseline stereo, IEEE Trans. on Pattern Analysis and Machine Intelligence, 15(4), 1993, s.353-363.

APPLICATIONS OF THE MSA ALGORITHM (MULTIPLE SIMILAR AREAS) FOR OBTAINING DEPTH MAPS BY USING MULTI-VIEW COMPUTER VISION SYSTEMS

The paper described a method for obtaining depth maps on the basis of images from a set of multiple cameras with the use of stereoscopic vision. Depth maps contain data about distances from objects located in the field of view of camera. A depth map can be acquired from images from a pair of cameras which has a function of a stereo camera. However, depth maps obtained by this method contain errors. This paper proposes a method of reducing errors by increasing the number of cameras. The paper presents the usage of the MSA algorithm (Multiple Similar Areas) designed for a set of cameras arranged in the configuration called EBMCS (Equal Baseline Multiple Camera Set). The configuration consists of a central camera and multiple side cameras. Obtaining 3D images with the use of multi-camera systems is an alternative to using other kind of devices such as structured light scanners and equipment based on LIDAR (Light Detection and Ranging).

Keywords: depth map, stereoscopic vision, multi-camera systems, camera array.