

## PORÓWNYWANIE ZDJĘĆ ROŚLIN NA PODSTAWIE MIAR PODOBIENSTWA RÓŻNIĄCYCH SIĘ WYKŁADNIKIEM POTĘGI

Adam L. KACZMAREK

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

**Streszczenie:** Artykuł przedstawia studium przypadku poświęcone zastosowaniu różnych miar podobieństwa do porównywania zdjęć roślin. Miary te mają zastosowanie w przypadku użycia układu kamer składającej się z pięciu kamer umieszczonych w bliskiej odległości. Badany układ kamer, nazwany *EBCA* (*Equal Baseline Camera Array*), składa się z jednej kamery centralnej oraz kamer bocznych. Opisane miary podobieństwa stosuje się w algorytmach widzenia stereoskopowego pozwalających na oszacowanie odległości między kamerami a obiektami widocznymi na zdjęciach. W artykule zaproponowane zostało uogólnienie stosowanych dotychczas miar tj. *SAD* (*Sum of Absolute Differences*) i *SSD* (*Sum of Squared Differences*). Przeprowadzone eksperymenty świadczą o tym, że zaproponowane miary pozwalają na redukcję błędów oszacowania polegających na otrzymaniu wyników odbiegających od prawidłowych wartości o przyjętą wartość progową.

**Słowa kluczowe:** widzenie stereoskopowe, łańcuch kamer, mapa rozbieżności, mapa głębi.

### 1. WSTĘP

Artykuł poświęcony jest tematowi widzenia stereoskopowego [1]. Widzenie stereoskopowe polega na użyciu pary kamer lub innych urządzeń rejestrujących obraz do celu określania odległości. Odległość od obiektów rzeczywistych będących w zasięgu widzenia określana jest na podstawie analizy różnic w zawartości zdjęć. Dokładność oszacowania odległości zależy jest między innymi od jakości algorytmów analizujących zdjęcia.

W celu poprawienia dokładności pomiaru możliwe jest również użycie większej liczby kamer. Niniejszy artykuł powstał w ramach prac poświęconych układowi kamer o nazwie *EBCA* (*Equal Baseline Camera Array*) [2,3,4].

### 2. KAMERY STERESKOPOWE

W kamerze stereoskopowej występuje kamera referencyjna oraz kamera boczna. Kamera referencyjna jest tą kamerą, względem której określana jest odległość od pary kamer do obiektów będących w zasięgu ich widzenia. Kamera referencyjna stanowi ponadto punkt widzenia układu. Część obiektów może być częściowo zasłonięta przez inne obiekty znajdujące się bliżej kamer. Niektóre fragmenty obiektów są widoczne tylko z kamery referencyjnej lub tylko z kamery bocznej.

Parę zdjęć pozyskaną z kamery stereoskopowej stosuje się do tego, aby wydobyć informacje o odległości kamery od widocznych na zdjęciach obiektów. W tym celu analizuje się położenie obiektów na zdjęciach. Jeśli jakiś obiekt jest

bardzo odległy od pary kamer wtedy w przypadku każdego z dwóch zdjęć wykonanych różnymi kamerami położenie obiektu na zdjęciach będzie prawie takie samo. Położenie to wyznaczone jest względem brzegów zdjęć. W przypadku gdy obiekt znajduje się blisko kamer jego położenie na lewym zdjęciu będzie znacząco różnić się od położenia na prawym zdjęciu.

Na podstawie pary zdjęć tworzona jest mapa rozbieżności (ang. *disparity map*). Mapa ta odzwierciedla różnice w lokalizacji obiektów na różnych zdjęciach. W celu wydobywania mapy rozbieżności stosuje się algorytmy dopasowujące fragmenty zdjęć (ang. *stereo matching algorithms*), których celem jest zidentyfikowanie w parze zdjęć tych samych fragmentów obiektów rzeczywistych. W wyniku działania algorytmów dopasowujących mogą wystąpić błędy polegające na tym, że nieprawidłowo zinterpretowane zostało to, że pewne dwa fragmenty różnych zdjęć przedstawiają ten sam fragment obiektu rzeczywistego, podczas gdy fragmenty te, mimo że są do siebie podobne przedstawiają części różnych obiektów.

Podczas porównywania fragmentów par zdjęć stosuje się pewne miary podobieństwa. Do najpopularniejszych z nich należy miara *SAD* (*Sum of Absolute Differences*) i *SSD* (*Sum of Squared Differences*) podane wzorami (1) [5]. Określanie podobieństwa dwóch punktów przy użyciu tych miar polega na określaniu podobieństwa fragmentów obrazów położonych dookoła badanych punktów. Fragmenty te nazywane są oknami agregującymi. Typowe okno agregujące ma kształt kwadratu, w którego środku znajduje się punkt, dla którego mierzone jest podobieństwo.

$$SAD(p, d) = \sum_{p \in W} |I_R(p) - I_S(p - d)| \quad (1)$$

$$SSD(p, d) = \sum_{p \in W} (I_R(p) - I_S(p - d))^2$$

gdzie:  $p$  – współrzędne punktu,  $d$  – wartość rozbieżności,  $I_R$  – wartość na zdjęciu z kamery referencyjnej,  $I_S$  – wartość na zdjęciu z kamery bocznej,  $W$  – punkty wchodzące w skład okna agregującego

Opracowanych zostało również wiele innych miar podobieństwa, między innymi *NCC* (*normalized cross correlation*) oraz *SSIM* (*structural similarity*), jednak powszechność ich stosowania w stereoskopii jest stosunkowo niewielka w porównaniu z popularnością miar *SAD* i *SSD*. Oprócz miar podobieństwa w algorytmach dopasowujących stosuje się inne metody mające na celu

zmniejszenie liczby błędów dopasowania. Do metod tych należy między innymi optymalizacja globalna lub identyfikowanie punktów charakterystycznych [1, 6]. Ten artykuł poświęcony jest jednak tematowi miar podobieństwa fragmentów obrazów. Miary te stosowane są w algorytmach dopasowujących niezależnie od tego, jakie metody analizy zdjęć zostały zastosowane w algorytmie dopasowującym.

Z mapy rozbieżności można otrzymać mapę głębi (ang. *depth map*) przedstawiającą odległości pomiędzy kamerą referencyjną, a obiektami. Wygenerowanie takiej mapy wymaga posiadania dodatkowych danych o parze kamer. Do danych tych należy między innymi odległość między kamerami oraz ogniskowe obiektywów.

### 3. UKŁAD KAMER EBCA

Badania przedstawione w tym artykule poświęcone są układowi kamer o nazwie *EBCA* (*Equal Baseline Camera Array*). Układ ten nazywany był wcześniej *Equal Baseline Multiple Camera Set (EBMCS)*, jednak obecna nazwa dokładniej przedstawia jego cechy, co ułatwi innym osobom znajdowanie informacji o tym układzie. Pierwszymi badaczami, którzy zaproponowali używanie tego układu kamer byli Park i Inoue [7]. Autor niniejszego artykułu kontynuuje badanie możliwości układu.

*EBCA* jest układem kamer, w którym występuje kamera centralna oraz kamery boczne. *EBCA* ma wiele cech wspólnych z kamerą stereoskopową. Kamerą referencyjną w układzie *EBCA* jest zawsze kamera środkowa. Układ ten można postrzegać jako zbiór czterech połączonych ze sobą kamer stereoskopowych współdzielących kamerę referencyjną. Rzeczywisty układ *EBCA* użyty w badaniach przedstawiony jest na rysunku 1. Urządzenie to składa się z pięciu kamer MS LifeCam Studio wyposażonych w sensor HD o rozdzielczości 1080p. Kamery zostały połączone aluminiowymi listwami.



Rys. 1. Układ kamer o nazwie EBMA (Equal Baseline Camera Array) użyty w eksperymentach

Stosowanie *EBCA* do widzenia stereoskopowego można zinterpretować jako wykonywanie czterech pomiarów czterema składającymi się na *EBCA* kamerami stereoskopowymi. Dzięki otrzymywaniu map rozbieżności mając do dyspozycji różne kamery boczne możliwe jest uzyskanie dokładniejszych wyników niż wtedy, gdy stosuje się pojedynczą kamerę stereoskopową.

Istotną zaletą układu *EBCA* jest to, że kamera referencyjna jest ta sama w każdej rozpatrywanej parze kamer. Pozwala to w prosty sposób scalać dane pozyskane przy użyciu różnych kamer bocznych. Układ *EBCA* ma ponadto inną bardzo istotną zaletę, mianowicie taką, że odległości między kamerami w analizowanych parach kamer są takie same. Dzięki temu mapy rozbieżności odwzajemniają te same wartości przesunąć niezależnie

od tego, które zdjęcie boczne zostało użyte do utworzenia mapy.

### 4. MIARY PODOBIENSTWA

Podobnie jak w przypadku analizowania zdjęć z kamer stereoskopowych, w przypadku analizy zdjęć z układu *EBCA* konieczne jest zastosowanie pewnej miary podobieństwa pozwalającej na zidentyfikowanie fragmentów zdjęć przedstawiających te same części obiektów rzeczywistych. Wzory (2) przedstawiają adaptację stosowanych z kamerami stereoskopowymi wzorów (1) do układu *EBCA*. Dostosowanie tych wzorów do układu *EBCA* polegająca na obliczaniu sumy wartości *SAD* i *SSD* dla wszystkich czterech kamer stereoskopowych branych pod uwagę w układzie *EBCA*.

$$SSAD(p, d) = \sum_{j=1}^N SAD_j(p, d) \quad (2)$$

$$SSSD(p, d) = \sum_{j=1}^N SSD_j(p, d)$$

gdzie: *N* – liczba kamer bocznych w układzie, *j* – indeks kamery bocznej, *SSD* i *SAD* oznacza miary podane równaniami (1).

Inna miarę podobieństwa, która może być stosowana w układzie kamer *EBCA*, zaproponowana została przez autorów Park i Inoue [7]. Rozróżniali oni pionowe i poziome pary kamer, a w zaproponowanym przez nich wzorze znajdowana była minimalna wartość wśród każdej z tych par. Wzór omawianych autorów oznaczony jako *PiI* podany jest równaniem (3).

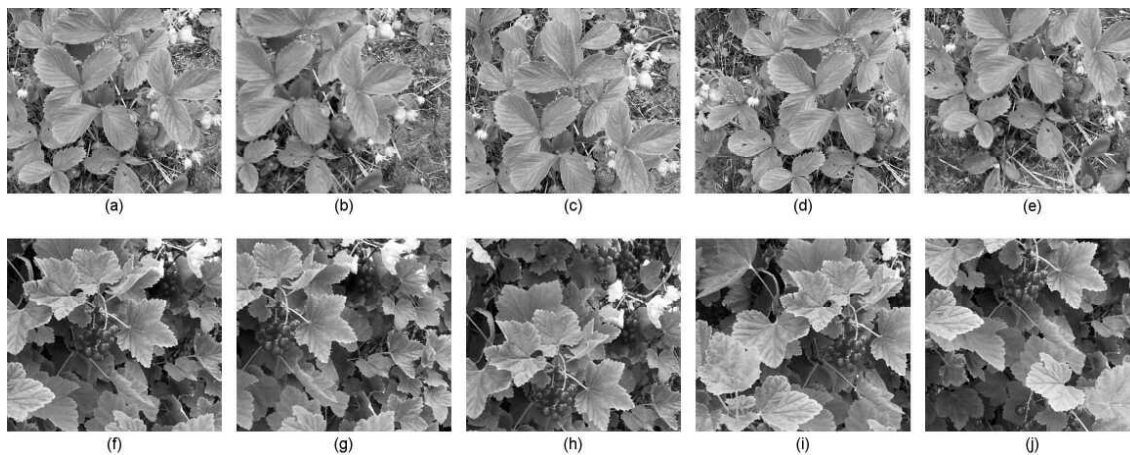
$$PiI(p, d) = \min(SSD_1, SSD_3) + \min(SSD_2, SSD_4) \quad (3)$$

gdzie: indeks parzysty podany przy wartości *SSD* odnosi się do par kamer położonych w osi pionowej, indeks nieparzysty dotyczy par kamer położonych w osi poziomej.

Ten artykuł przedstawia eksperymenty poświęcone badaniu uogólnienia wzorów (2) oraz (3). Wzory (1) na *SAD* i *SSD* można zapisać w postaci uogólnionej wzoru *SPD* (*Sum of Power Differences*) przedstawionego równaniem (4). Wzór (4) można wówczas użyć zamiast użytych we wzorach (2) i (3) funkcji *SAD* i *SSD*. Wzory *SSAD* i *SSSD* oznaczone zostaną wówczas jako wzór *SSPD* natomiast wzór *PiI* oznaczony zostanie jako *PiIP*.

$$SPD(p, d) = \sum_{p \in W} |I_R(p) - I_S(p - d)|^\alpha \quad (4)$$

Współczynnik  $\alpha$  jest potęgą, do której podnoszona jest różnica wartości w oknach agregujących dwóch obrazów. Gdy  $\alpha=1$  wzór przyjmuje postać wzoru *SAD*, natomiast dla  $\alpha=2$  jest to wzór przedstawiający wartość *SSD*. Wartości naturalne współczynnika  $\alpha$  stosuje się w celu przyspieszenia obliczeń. Jednak mając na uwadze dużą moc obliczeniową współczesnych komputerów autor sprawdził, jakie są efekty zastosowania współczynnika  $\alpha$  o wartościach wymiernych w zakresie od 0 do 2. Wyniki przedstawione są w rozdziale 6. Użyte zostało okno agregujące o rozmiarze 9×9. Jest to okno o stosunkowo dużym rozmiarze powodujące podniesienie jakości wyników kosztem znaczącej liczby obliczeń niezbędnych do przeprowadzenia podczas przetwarzania danych zawartych w oknie.



Rys. 2. Przykładowe zestawy zdjęć truskawki ((a)-(e)) i porzeczki ((f)-(j)) użyte w eksperymentach. W przedstawionej kolejności są to zdjęcia z kamery środkowej, prawej, górnej, lewej i dolnej [4]

## 5. DANE EKSPERYMENTALNE

Eksperymenty przeprowadzone zostały w oparciu o dane testowe przedstawione w poprzednich pracach poświęconych układowi *EBCA* [2][3]. Dane testowe składają się z sześciu zestawów zdjęć wykonanych badanym układem. Każdy zestaw zawiera zdjęcia utworzone różnymi kamerami wchodzącymi w skład układu. Obiektami fotografowanymi były rośliny z dojrzałymi owocami truskawki, porzeczki oraz wiśni. Utworzone zostały po dwa zestawy danych dla każdej z tych roślin. Rośliny zostały wybrane do testów ze względu na ich złożony kształt powodujący, że wykonywanie zdjęć i skanów trójwymiarowych tego rodzaju obiektów jest problematyczne. Dane testowe zawierają ponadto prawidłowe wartości rozbieżności, które powinny zostać otrzymane w wyniku działania algorytmu dopasowującego. Opis sposobu pozyskania tych danych przedstawiony został w [3].

Przykładowe zestawy zdjęć pokazane są na rysunku 2. Zdjęcia te zostały przetworzone w procesie kalibracji, która spowodowała, że zredukowane zostały konsekwencje tego, że występują niedokładności związane z mocowaniem kamer w rzeczywistym układzie *EBCA*. W wyniku występowania tego rodzaju niedokładności osie optyczne kamer nie są w stosunku do siebie równoległe a kamery są obrócone względem siebie. Kalibracja danych użytych w eksperymentach opisana dostała w [3]. Zdjęcia testowe różniły się rozmiarem. Sumaryczna liczba wszystkich punktów testowych równa była 172400.

W celu ewaluacji wyników map rozbieżności konieczne jest przyjęcie pewnej miary jakości. W tym artykule wyniki opierają się na powszechnie stosowanej mierze *BMP* (ang. *bad matching pixels*) [1]. Miara ta podaje odsetek punktów mapy rozbieżności, dla których wartość otrzymana algorytmem dopasowującym różni się od wartości prawidłowej o daną wartość progową  $Z$ . Wartości  $Z$  mogą być różne w zależności od wymagań na dokładność mapy rozbieżności. Dla wyników przedstawionych w tym artykule zostało przyjęte, że wartość progowa jest równa 2. Przyjęcie innego progu zmieniłoby wartości *BMP* dla wszystkich zdjęć testowych. Jednak podczas ewaluacji istotne jest przede wszystkim porównanie

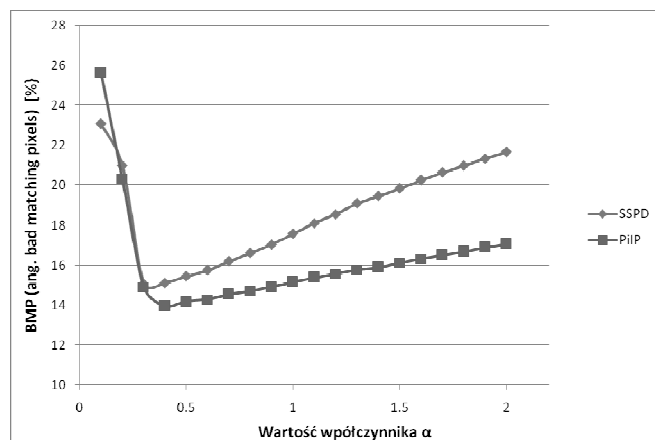
wyników rodzaju miar podobieństwa stosując tą samą miarę jakości. Zastosowanie innej wartości progowej jest również możliwe. Miara *BMP* przedstawiona jest wzorem 5.

$$BMP = \frac{1}{N} \sum_x (|D_M(x) - D_T(x)| > Z) \quad (5)$$

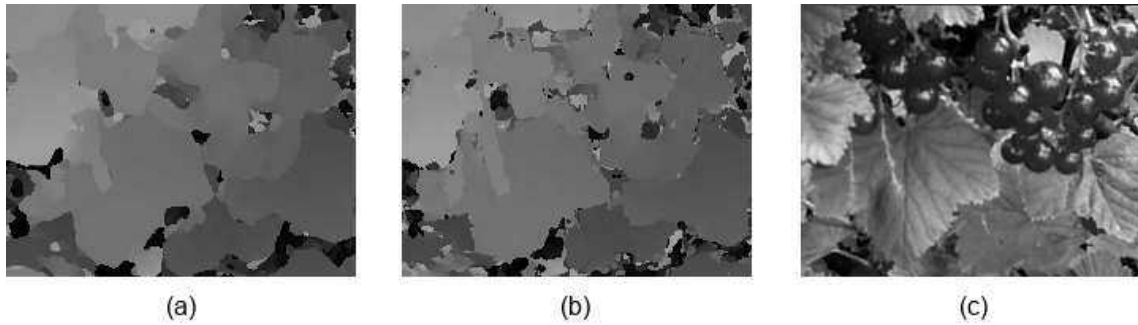
gdzie:  $x$  – współrzędne punktu,  $D_M(x)$  – wartość na mapie rozbieżności,  $D_T(x)$  – wartość rzeczywista,  $N$  – liczba punktów,  $Z$  – wartość progowa.

## 6. WYNIKI

Wyniki eksperymentów przedstawione są na rysunku 3. Wykres przedstawiający wyniki zastosowania miar podobieństwa *SSPD* i *PiIP* powstał dla współczynnika  $\alpha$  w zakresie od 0,1 do 2,0 z krokiem 0,1. Mniejsze wartości *BMP* świadczą o wyższej jakości mapy rozbieżności. W związku z tym najlepsze rezultaty osiągnane są wtedy, gdy użyta zostanie miara *PiIP* ze współczynnikiem  $\alpha$  równym 0,4. Ponadto wykres wskazuje na to, że zastosowanie miary *PiIP* w ogólności prowadzi do lepszych wyników niż użycie miary *SSPD* dla tych samych wartości  $\alpha$ .



Rys. 3. Wartości miary *BMP* otrzymane dla miar *SSPD* i *PiIP*



Rys. 4. Przykładowe fragmenty map rozbieżności otrzymane w wyniku wykonania eksperymentów dla następujących miar: (a) *SSPD*,  $\alpha=0,4$ , (b) *PiIP*,  $\alpha=2$ , (c) fragment zdjęcia odpowiadający mapom

Przykładowe fragmenty map rozbieżności uzyskane opisanymi wyżej sposobami przedstawione są na rysunku 4. Przedstawione na tym rysunku mapy przedstawiają różnym poziomem jasności różne wartości rozbieżności. Im jaśniejszy jest punkt tym większa jest rozbieżność. Część (a) przedstawia mapę rozbieżności uzyskaną dla miary *SSPD* i współczynnika  $\alpha$  równego 0,4. Część (b) wizualizuje mapę dla miary *PiIP* i  $\alpha=2$ . Część (c) przedstawia fragment zdjęcia odpowiadający pokazanym mapom.

## 7. PODSUMOWANIE

Badania opisane w tym artykule pozwalają podnieść jakość map rozbieżności dzięki zastosowaniu lepszych miar podobieństwa. W artykule przedstawione zostały niepublikowane wcześniej wyniki eksperymentów. W powszechnie stosowanych algorytmach użycie opisanych wyżej miar stanowi jeden z wielu kroków składających się na wynik końcowy. W dalszych badaniach planowane jest zastosowanie przedstawionych miar do złożonych algorytmów widzenia stereoskopowego, takich jak algorytm *StereoSGBM* (*Stereo Semi-Global Block Matching*) dostępny w bibliotece programistycznej *OpenCV* [8].

## 8. 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. 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.
3. Kaczmarek A. L.: Stereo vision with equal baseline multiple camera set (EBMCS) for obtaining depth maps of plants, *Computers and Electronics in Agriculture* 135, 2017, s. 23-37.
4. Kaczmarek A. L.: Wpływ nadmiarowych danych wejściowych na jakość map rozbieżności pozyskanych za pomocą układu pięciu kamer w konfiguracji EBMCS, *Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej* 60, Gdańsk 2018, s.41-44.
5. 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.
6. Geiger A., Roser M., Urtasun R.: Efficient large-scale stereo matching, *Computer Vision – ACCV 2010, Lecture Notes in Computer Science*, vol. 6492, Springer, 2011, s. 25-38.
7. Park J., Inoue S.: Acquisition of sharp depth map from multiple cameras, *Signal Processing: Image Communication*, 14 (1-2), 1998, s. 7-19.
8. Bradski D.G.R., Kaehler A.: *Learning OpenCV*, 1st Edition., O'Reilly Media Inc., 2008.

## COMPARISON OF IMAGES OF PLANTS ON THE BASIS OF SIMILARITY METRICS WITH DIFFERENT EXPONENTS

The paper presents a case study concerned with applying different similarity measures for comparing images of plants. These measures are used for a camera array which consists of five adjacent cameras. The researched array called Equal Baseline Camera Array (*EBCA*) contains one central camera and four side cameras. The described measures can be used with stereo vision algorithms designed for estimating distances between cameras and objects visible in images taken with the use of these cameras. The paper introduces generalizations of currently used measures such as Sum of Absolute Differences (*SAD*) and Sum of Squared Differences (*SSD*). The experiments show that the proposed measures make it possible to reduce estimation errors which occur in results differing from right values more than a selected threshold.

**Keywords:** stereo vision, camera array, disparity map, depth map.