Computers and Electronics in Agriculture 135 (2017) 23–37

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

# Stereo vision with Equal Baseline Multiple Camera Set (EBMCS) for obtaining depth maps of plants

Adam L. Kaczmarek

*Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, ul. G. Narutowicza 11/12, 80-233 Gdansk, Poland*

ARTICLE INFO

*Article history:*
Received 26 April 2016
Received in revised form 16 August 2016
Accepted 25 November 2016
Available online 7 February 2017

*Keywords:*
Depth maps
Stereo matching
Camera matrix
Camera array
Robotic harvesting

ABSTRACT

This paper presents a method of improving the estimation of distances between an autonomous harvesting robot and plants with ripe fruits by using the vision system based on five cameras. The system is called Equal Baseline Multiple Camera Set (EBMCS). EBMCS has some features of a camera matrix and a camera array. EBMCS is regarded as a set of stereo cameras for estimating distances by obtaining disparity maps and depth maps. This paper introduces Exceptions Excluding Merging Method (EEMM) which makes it possible to improve the quality of disparity maps by integrating maps acquired from individual stereo cameras included in EBMCS. The method was tested with eight different stereo matching algorithms including Efficient Large-scale Stereo Matching (ELAS), algorithms implemented in the OpenCV library and algorithms available in Middlebury Stereo Vision Page. Experiments were performed on input data sets which contained images of strawberry plants, cherry trees and redcurrant plants. The bad matching pixels (BMP) metric was used for measuring the error rate in disparity maps used in the distance estimation. The results of experiments showed that, on average, the EEMM merging method used with EBMCS consisting of five cameras reduces the error rate of the distance estimation by 26.55% in comparison to results obtained from stereoscopy based on a single stereo camera. The best results were acquired by using with five cameras a stereo matching algorithm based on a graph cut.



## 1. Introduction

Many fruit farmers anticipate advances in the development of robotic fruit harvesting. This technology has a potential to revolutionize the process of harvesting corps, however it is still rarely used in the field. Fruits can be picked up by autonomous robots which are not directly controlled by human beings.

One of the most important elements of an autonomous robot designed for picking up fruits is the vision system. The purpose of the vision system is both to recognize fruits and to estimate distances between fruits and the robot. The recognition of fruits is based on cameras installed in robots (P. Li et al., 2011). The estimation of distances is performed with the use of different devices including cameras (van Henten et al., 2002; Xiang et al., 2010; Hayashi et al., 2010), TOF (time-of-flight) cameras (Kazmi et al., 2012) and structured-light 3D scanners (Jang et al., 2013). This paper proposes estimating distances by using a set of multiple cameras which has a better performance than devices listed above.

TOF cameras take advantage of data about the speed of light. These devices emit rays of light and records reflection of these rays from the analyzed objects. The distance is determined on the basis of the time needed by the light to reach the object and to return to the measuring device afterward.

Structured-light 3D scanners illuminate analyzed objects with light shaped into strictly defined patterns (Jang et al., 2013). For example, the light pattern can have a form of parallel stripes. The scanner records distortions of the light on the analyzed objects. On this basis, distances are determined.

The results of TOF cameras and structured-light scanners are very accurate, but these devices have a major disadvantage – they emit light in order to perform the measurement. Natural light interfere with the light emitted by the devices disrupting their operations. Consequently, the quality of results deteriorates when analyzed objects are exposed to intensive natural light that occurs in plant fields (Kazmi et al., 2012, 2014; Gupta et al., 2013).

In this paper, camera based distance estimation is used, which profits from high level of illumination. A distance between a viewpoint and observed objects can be determined to some extent on the basis of sizes of objects on a single image taken from only one camera (Baeten et al., 2008). However, commonly used methods of estimating distances are based on a stereo camera that is a set of two cameras located at the side of each other. Both cameras in the set are pointed in the same direction. Stereo matching

*E-mail address:* adam.l.kaczmarek@eti.pg.gda.pl
*URL:* http://pg.edu.pl/d119a02021_adam.kaczmarek

http://dx.doi.org/10.1016/j.compag.2016.11.022
0168-1699/

algorithms designed for stereo cameras estimate distances to viewed object by the analysis of a pair of images taken from different viewpoints. The estimation makes it possible to obtain a depth map which is a set of distances between a viewpoint and viewed objects. In order to improve the precision of distance estimation, a greater number of cameras can be used (Okutomi and Kanade, 1993; Nielsen et al., 2007; Hensler et al., 2011).

The research on determining distances on the basis of multi-camera vision systems is very limited in the field of agriculture. However, the agriculture and the out-door environment is particularly suitable for this kind of equipment. This paper contributes to this research area. The paper presents research on taking advantage of a set of cameras in order to estimate distances to plants and their parts. The number of cameras in considered sets ranges from two to five. Cameras are arranged in a specific form which is described in Section 3.1 of this paper. The applied camera arrangement has been called by the author Equal Baseline Multiple Camera Set (EBMCS). Such a set of cameras is intended for use with a robotic arm of an autonomous robot designed to harvest fruits.

The research presented in this paper is based on a previous research performed by the author (Kaczmarek, 2015). Both papers refer to the same kind of a camera set. However, the previous research was focused on developing an algorithm for obtaining depth maps which was dedicated for the considered set of cameras. This paper presents methods for applying any kind of a stereo camera depth map making algorithm to the set of cameras arranged as EBMCS.

The original contributions of this paper are the following: (1) The design of a novel method for determining distances called Exceptions Excluding Merging Method (EEMM). It is intended for use with Equal Baseline Multiple Camera Set (EBMCS). On average, EEMM reduces the error rate of the distance estimation by over 26% when five cameras are used instead of two ones; (2) The analysis of the influence of the number of stereo cameras included in EBMCS on the quality of depth maps obtained with the use of this set; (3) The development of test data sets consisting of images of plants for testing of depth maps merging methods used with EBMCS.

## 2. Related work

Various areas of the robotic fruit harvesting technology have been researched, including the construction of an arm for picking fruits (Tanigaki et al., 2008), robotic arms control systems (Mehta and Burks, 2014), the navigation of the robot through an orchard (Murakami et al., 2008), fruit recognition algorithms (Hayashi et al., 2010; Bac et al., 2013) and vision systems for harvesting robots (Belforte et al., 2006). This paper focuses on vision systems.

### 2.1. Vision systems in robotic fruit harvesting

Li, Lee and Hsu presented a review on the technology of harvesting citrus fruits (P. Li et al., 2011). They focused on citrus fruits because they noticed that harvesting mechanisms are more advanced for this kind of fruits than for other ones. Belforte et al., presented a general review on a harvesting technology (Belforte et al., 2006). Grift et al. prepared a very detailed review on automation methods and robotics for the bioindustry including robotic harvesting systems (Grift et al., 2008).

Harvesting robots take advantage of different kinds of vision systems in order to determine distances to fruits. Some devices for automatic fruit harvesting use only one camera to estimate the distance. The measurement is based on the size of a fruit in the image made by a camera. This method was used by Baeten et al. in automatic apple harvesting (Baeten et al., 2008). A single camera has also been used by Muscato et al. in a robot designed to harvest oranges (Muscato et al., 2005).

More precise methods of estimating distances to objects are vision systems based on stereo cameras. In such a pair of cameras there is a reference camera and a side camera. A stereo matching algorithm matches patterns visible in an image from the reference camera with corresponding patterns in an image from the side camera. The difference in the location of the same pattern in these two images is a disparity. The set of disparities concerning the same image forms a disparity map. A disparity map can be converted to a depth map. The value of depth is explicitly defined by the value of disparity. Obtaining depths from disparities requires collecting data about stereo camera parameters such as the distance between cameras and focal lengths of lens (Okutomi and Kanade, 1993).

Van Henten et al. constructed an autonomous robot for harvesting cucumbers equipped with two cameras (van Henten et al., 2002). A manipulator and an end-effector of the robot picked up cucumbers located by cameras in the 3D space. Harvesting devices were placed on an autonomous vehicle which moved along aisles of a greenhouse. Xiang et al. used a stereo vision system for locating tomatoes (Xiang et al., 2010). They have analyzed the relation between distances from cameras to tomatoes and the accuracy of disparity maps obtained for these fruits.

Hayashi et al. applied a stereo camera to a robot for picking up strawberries (Hayashi et al., 2010). The robot took advantage of three cameras. However, only two of them were used for obtaining depth maps. Hayashi et al. deployed a fully equipped robot which makes is possible to harvest strawberries without direct human control. Stereo cameras were also used by Plebe and Grasso in a robot designed for picking oranges (Plebe and Grasso, 2001). The device had two telescopic arms. Each one of them contained a stereo camera. Stereo cameras obtained depth maps independently of each other.

Autonomous harvesting robots were also equipped with stereo cameras delivered in the form of a single device. This kind of a product is Point Grey BumbleBee2. Qingchun et al. used this device in a robot for harvesting strawberries (Qingchun et al., 2012). The stereo camera has been located on a robotic arm designed for picking up fruits. Point Grey BumbleBee2 has also been used by Yang et al. in a fruit recognition system for harvesting tomatoes (Yang et al., 2007).

Depth maps of plants were also obtained with the use of three-camera vision systems. Such a set was used by Nielsen et al. (2007). They compared the performance of stereo matching algorithms applied to images of real plants and images of plants artificially rendered by computers. In their research they focused on the stereo matching algorithm which used the Sum of Squared Difference measure with Symmetric Multiple Windows (SMW) (Fusiello et al., 2000).

There has also been research on using making 3D models of plants on the basis of images taken from different, overlapping points of views. In this process, a structure from motion is obtained. Tan et al. presented results of this kind of modeling applied to trees (Tan et al., 2007). They obtained 3D models by using over 10 images taken from different locations placed around these plants. The similar kind of a research was presented by Quan et al. for plants growing in pots (Quan et al., 2006).

Depth maps of plants can also be obtained with the use of devices other than cameras. Chéné et al. used the Microsoft Kinect camera for making depth maps of plants leaves (Chéné et al., 2012). The device that they have used was a kind of a structured-light 3D scanner.

Tanigaki et al. presented a robot for harvesting cherries which estimated distances to parts of plants by using a laser beam (Tanigaki et al., 2008). The disadvantage of using laser beams is

such that a single measurement provides information about the distance to a single point. In order to cover a larger number of points a series of measurements need to be performed. This problem does not occur when a TOF camera is used. This device is also based on a laser, but is discovers distances to many points with a single execution. Kazmi et al. analyzed the influence of an intensive natural light on the results of a TOF camera (Kazmi et al., 2012, 2014). They have performed experiments on leaves of plants. In general, an increase in the illumination has a negative impact on the performance of a TOF camera. A multi-camera system proposed in this paper is an alternative to the technology currently used in the field of robotic harvesting.

### 2.2. Stereo camera vision systems

Disparity maps of plants can be obtained by one of many general purpose stereo matching algorithms. Scharstein, Szeliski and Hirschmller provided a ranking of such algorithms in Middlebury Stereo Vision Page (http://vision.middlebury.edu/stereo/) (Scharstein and Szeliski, 2002; Hirschmuller and Scharstein, 2007). The version 2 of the ranking considers over 160 different stereo matching algorithms. Algorithms included in this list were used in experiments presented in this paper.

The ranking prepared by Szeliski is based on a testbed consisting of exemplary images from stereo cameras. Each set of images included in the testbed also contains ground truth that is a map with real values of disparities. Every point of a ground truth map corresponds to a point of an image for which ground truth was prepared. Ground truth is used for estimating the quality of stereo matching algorithms. Middlebury Stereo Vision Page contains over 70 collections of images with ground truth.

Moreover, Middlebury Stereo Vision Page provides implementations of the following stereo matching algorithms (Szeliski et al., 2008): Iterated Conditional Modes (ICM) (Besag, 1986), Graph Cuts using Swap Moves (GC Swap) (Boykov et al., 2001), Graph Cuts using Expansion Moves (GC Expansion) (Boykov et al., 2001), Max-product Loopy Belief Propagation (LBP) (Tappen and Freeman, 2003) and Sequential Tree-reweighted Message Passing (TRW-S) (Wainwright et al., 2005; Kolmogorov, 2006). These algorithms compute disparity maps by minimizing an energy function which depends on a distribution of Markov Random Fields (MRF) (Besag, 1986). The purpose is to reduce inconsistencies in disparity maps and to enforce spatial coherence. Algorithms GC Swap, GC Expansion, BP-M (the version M of the LBP algorithm) and TRW-S were used in experiments with multi-camera vision systems presented in this paper (Section 6).

ICM is the oldest method from those implemented in Middlebury Stereo Vision Page. However, it lacks effectiveness and efficiency (Besag, 1986; Szeliski et al., 2008). Middlebury Stereo Vision Page also provides implementations of two algorithms based on graph cuts. They are called Swap-move and Expansion-move (Kolmogorov and Zabin, 2004; Boykov and Kolmogorov, 2004). Both of these algorithms iteratively converge to a global minimum i.e. the minimal value of an energy function. The Swap-move algorithm requires switching values in different graph nodes. The Expansion-move method modifies values without this requirement.

Belief Propagation is an algorithm used in different domains. Its version called Max-Product Loopy Belief Propagation has been applied for computing disparity maps (Tappen and Freeman, 2003). The algorithm is based on passing messages along nodes. Middlebury Stereo Vision Page includes two implementations of this algorithm: one denoted BP-M and the other denoted BP-S. The order of data processing is one of the main differences between them. Middlebury Stereo Vision Page also provides the sequential Tree-Reweighted Message Passing (TRW-S) algorithm (Wainwright et al., 2005; Kolmogorov, 2006). The algorithm is similar to the Loopy Belief Propagation algorithm.

Apart from Middlebury Stereo Vision Page the OpenCV library (http://opencv.org/) is another key project in the field of stereo vision. The library provides various functions for computer vision and 4 stereo matching algorithms (Bradski and Kaehler, 2008). These algorithms are denoted by StereoBM (Stereo Block Matching) (Konolige, 1998), StereoSGBM (Stereo Semi-Global Block Matching) (Hirschmuller, 2008), StereoHH (Heiko Hirschmüller algorithm) (Hirschmuller, 2008) and StereoVar (Stereo Variational methods) (Kosov et al., 2009). Names are derived from names of classes and functions implemented in the library. These algorithms were also used in the experiments described in Section 6 of this paper similarly to algorithms available in Middlebury Stereo Vision Page.

StereoBM is a stereo correspondence block matching algorithm similar to the one developed by Bradski and Kaehler (2008) and Konolige (1998). The algorithm represents classic approach to matching points between left and right image on the basis of the sum of absolute differences (SAD). The OpenCV implementation includes pre-filtering and post-filtering such as uniqueness check, quadratic interpolation and speckle filtering (Bradski and Kaehler, 2008).

StereoSGBM and StereoHH are two versions of the same matching algorithm proposed by Hirschmuller (2008). There are some differences between the OpenCV implementation and the original algorithm (Bradski and Kaehler, 2008). Differences include matching blocks instead of pixels and using the Birchfield-Tomasi subpixel metric instead of the mutual information function cost (Birchfield and Tomasi, 1998). StereoSGBM takes into account 5 directions used in the Hirschmüller algorithm, while StereoHH includes 8 directions. StereoSGBM is therefore faster than StereoHH.

StereoVar is a stereo matching method based on the algorithm described by Kosov et al. (2009). Variational methods focus on estimating optic flow which is a displacement field of corresponding pixels (Kosov et al., 2009). There are also some differences between the OpenCV version of the algorithm and the original one (Bradski and Kaehler, 2008). However, both methods optimize the energy function by iterative solvers.

Experiments presented in this paper also include the stereo matching algorithm called Efficient Large-scale Stereo Matching (ELAS) (Geiger et al., 2011). A characteristic feature of ELAS is such that the algorithm first computes disparities for a set of points that can be robustly matched due to their uniqueness. These points are called support points. Their disparities are calculated for a full disparity range. Remaining points are matched with regard to support points. The algorithm automatically determines the disparity range of a disparity map. Authors of the algorithm provides its implementation in the LIBELAS library (http://www.cvlibs.net/software/libelas/).

### 2.3. Multi-camera vision systems

Algorithms for obtaining depth maps on the basis of a pair of cameras can be applied for making depth maps with the use a multi-camera vision system. This paper contributes to the development of depth maps making methods for multi-camera systems. One of the most significant research papers dedicated to multi-camera stereo visions was written by Okutomi and Kanade (1993). The paper is concerned with estimating distances to objects on the basis of a camera array. The array is a sequence of cameras which are aimed in the same direction and located along a straight line. Distances between neighboring cameras were equal to each other. Cameras were identified by subsequent numbers starting with the index 0.

The camera set was regarded as a sequence of stereo cameras which consisted of camera 0 and camera $n$, where $1 \leqslant n \leqslant N$. The value $N$ is the index of the last camera in the array. The image from camera 0 was a reference one in every pair of images. A significant problem with using such a set of stereo cameras is the fact that they have different baselines, i.e. the distance between camera 0 and camera $n$ is different for different values of $n$. As a consequence, disparities of points from an image made by camera 0 depend on the stereo camera for which the disparity is obtained.

This fact complicates merging data from different stereo cameras in the set. Okutomi and Kanade resolved this problem by estimating distances with regard to values of inverse distances from camera 0 to points of viewed objects instead of using values of disparities. The distance between camera 0 and a viewed point is the same regardless of the stereo camera used in calculations.

Okutomi and Kanade based their algorithm for making depth maps on the Sum of Sum of Squared Differences (SSSD) measure. They have proposed an algorithm which uses an SSSD-in-inverse-distance matching cost function. In the algorithm, the inverse value of distance for each point of the reference image was calculated with regard to all other images made by a camera array. The algorithm was based on identifying the lowest value of SSSD-in-inverse-distance. The function is presented in Eq. (1).

$$SSSD(\mathbf{x},\zeta) = \sum_{1\leqslant i\leqslant N}\sum_{\mathbf{j}\in\mathbf{W}}(I_0(\mathbf{x}+\mathbf{j}) - I_i(\mathbf{x}+\mathbf{B_i}F\zeta+\mathbf{j}))^2 \quad (1)$$

where $\zeta$ is the inverse distance, $\mathbf{x}$ denotes coordinates of a point, $N$ is the number of cameras $I_i$ is the intensity of the point in the image from the camera $i$, $F$ is the focal length of the camera, $\mathbf{W}$ is an aggregating window consisting of points located in the vicinity of the point at coordinates $\mathbf{x}$ and $\mathbf{B_i}$ is the baseline which is the distance between the camera 0 and the camera $i$.

The matching cost function compares differences between an aggregating window in the reference image with corresponding windows in other images. Aggregating windows consist of the point for which the depth is calculated and neighboring points. A map with inverse distances obtained as the result of the algorithm contains values of inverse distances for which the matching cost function returns the lowest results.

Other vision systems containing a large number of cameras have also been developed. Wilburn et al. presented a multiple camera set which contained up to 100 cameras (Wilburn et al., 2005). Cameras were placed along parallel rows forming a matrix. Different arrangements were used including a matrix containing 8 rows with 12 cameras in a row. The matrix was only used in an in-door environment. The main purpose of using the set with such a large number of cameras was making high-resolution videos and videos of objects moving at high speed.

Multi-camera sets were also used for making 3D movies. Matusik and Pfister used a set of 16 cameras (Matusik and Pfister, 2004). Apart from capturing 3D movies their TV system made it possible to display captured scenes in real-time. They have also experimented with making 3D videos of fast moving objects. There are 3D videos which provides 3D images of a scene from different points of view (Domański et al., 2013). In some 3D video coding standards multiple views are represented as 2D images and respective depth maps.

Making 3D images of stationary objects does not require using a large number of cameras. Park and Inoue proposed a set of five cameras for making 3D images of real objects (Park and Inoue, 1998). The set contained a central camera and four cameras around the central one. The same kind of the camera arrangement used by Park and Inoue has been used in the research presented in this paper. The camera set is fully described in Section 3.1.

Park and Inoue regarded their cameras as a set of four stereo cameras. Each stereo camera consists of a central camera and one of side cameras. They did not need to use values of inverse distances instead of disparities, because in their camera set each considered stereo camera had the same baseline. Park and Inoue used this camera set for making disparity maps of immobile objects such as buildings. They used for obtaining disparities a matching cost function which was a modified version of the SSSD measure. The function that they have used in presented in Eq. (2).

$$\hat{d}(\mathbf{x}) = \underset{d}{argmin}(min[SSD_{left}(\mathbf{x},d), SSD_{right}(\mathbf{x},d)] + min[SSD_{top}(\mathbf{x},d), SSD_{bottom}(\mathbf{x},d)]) \quad (2)$$

where $\mathbf{x}$ are coordinates of a point, $d$ is a disparity, $\hat{d}$ is the disparity selected for the resulting disparity map and $SSD$ denotes the Sum of Squared Differences matching function presented in Eq. (3).

$$SSD_i(\mathbf{x},d) = \sum_{\mathbf{j}\in\mathbf{W}}(I_0(\mathbf{x}+\mathbf{j}) - I_i(\mathbf{x}+\mathbf{j}+\mathbf{d_i}))^2 \quad (3)$$

where $\mathbf{d_i}$ is the disparity in a stereo camera $i$ and other symbols are the same as in Eq. (1).

The algorithm proposed by Park and Inoue creates two groups of matching functions. The first group consists of functions used for the right and the left camera. The second group are functions for the top and the bottom camera. The algorithm selects the lower result of matching functions within each group. The disparity selected for a disparity map is calculated by finding the value of a disparity for which the sum of results from two groups returns the lowest value (Eq. (2)). The algorithm proposed by Park and Inoue was tested in the experiments presented in this paper.

A similar technique for obtaining disparities map on the basis of a multi-camera set has been proposed by Hensler et al. (2011). They performed the research on a set consisting of four cameras. Likewise the set proposed by Park and Inoue the set with four cameras consisted of a central camera and side cameras. The distance between the central camera and side ones was the same in every pair. Hensler et al. applied the set for making depth maps used in face recognition algorithms and a 3D reconstruction of faces.

Moreover, Williamson and Thorpe applied a trinocular vision system to automotive engineering (Williamson and Thorpe, 1999). They have developed a highway obstacle detection system which took advantage of depth maps obtained from a set of three cameras. The problem of obtaining depth maps from a three-camera vision system was also researched by Agrawal and Davis (2002).

### 2.4. Calibration and rectification

The development of a multiple vision system requires calibrating and rectifying cameras. The calibration is performed to reduce image distortions caused by imaging devices. Because of distortion, straight lines of real objects become bowed in the image made by a camera. Additionally, the rectification of cameras is concerned with a mutual location of cameras. In real multiple camera systems there is always some inaccuracy which causes that optical axes of cameras are not parallel to each other. Calibration and rectification are usually performed by making several images of a sample pattern such as a black-white chessboard. These images indicate transformations that are required for reducing distortions and inaccuracies in images.

Zhang described a very widely used calibrating method (Zhang, 2000). The method requires to take images of a planar pattern at a few (at least two) different orientations. The technique proposed by Zhang is easier in use than methods in which 3D patterns were necessary. Hartley introduced a commonly used method of

rectifying stereo images (Hartley, 1999). The method is applicable to both calibrated and uncalibrated pairs of images from a stereo camera.

Algorithms proposed by Zhang and Hartley have been implemented in the OpenCV library (Bradski and Kaehler, 2008). These implementations are intended for use with a single pair of cameras. However, Deng et al. used OpenCV for calibrating and rectifying images taken by a multi-view system (Deng et al., 2010). In their experiments, they used three cameras, but their method can be applied for sets containing any number of cameras. The OpenCV library has been also used for calibrating and rectifying the camera set presented in this paper. The application of OpenCV for this purpose is described in Section 3.2.

The calibration of a multi-camera system can be performed with the use of the same algorithms that are used for calibrating stereo cameras. Cameras are calibrated independently from each other. However, a camera cannot be rectified without taking into account the location and parameters of other cameras.

Khang and Ho proposed a method for rectifying camera arrays (Kang and Ho, 2011). In their technique, all images from cameras are transformed with respect to coordinates system of the image from the first camera in the array. This feature is characteristic for the method because a large number of rectification methods transform images from all camera. Another method of rectifying multi-view sets has been described by Yang et al. (2014a,b). They have researched an ergodic method for rectifying camera matrices and arrays. Sun presented research on rectifying trinocular vision systems (Sun, 2003). He proposed rectification methods for uncalibrated images.

The calibration and the rectification methods described above are concerned with geometric properties. However, a stereo camera and multi-camera vision systems require also the calibration of colors. Colors in images taken by different cameras are inconsistent because of fabrication variations of cameras and different light conditions caused by different camera locations (K. Li et al., 2011). In general, there are two methods for performing a color calibration of cameras. The first method is based on images statistics or values of colors in characteristic points of images. The other one is based on taking images of a color pattern.

Gurbuz et al. proposed a method for calibrating colors based on adjusting them with respect to color values in characteristic points of images (Gurbuz et al., 2010). They identified characteristic points with the use of the Scale-invariant Feature Transform (SIFT) algorithm. Nanda and Cutler took into account statistics of a whole image, for example they calculated a mean brightness (Nanda and Cutler, 2001). There are also methods that take advantage of exemplary patterns such as rectangular shapes painted in different colors. This kind of a calibration was described by K. Li et al. (2011) and Kurillo et al. (2013).

## 3. Equal Baseline Multiple Camera Set

This paper presents research on obtaining depth maps of plants with the use of four camera sets which differ in the number of cameras that they contain. The number of included cameras is in the range from two to five. The novel method of merging data proposed in this paper makes it possible to improve the quality of results by taking advantage of a larger number of cameras.

### 3.1. Arrangement of cameras

The set containing the greatest number of cameras consists of a central camera and four side cameras. It is the same kind of a set that used Park and Inoue (1998) and Kaczmarek (2015). All cameras were aimed in the same direction. Side cameras are located above, below and at both sides of the central camera. Therefore, there is right, up, left and down camera denoted with index 0, 1, 2, 3 and 4, respectively. The arrangement of cameras is presented in Fig. 1.

Every side camera creates a single stereo camera with the central camera. Stereo cameras will be marked with the same indexes as side cameras that are included in these stereo cameras. Distance between each side camera and the central one is equal for every side camera. Thus, the set of five cameras consists of four stereo cameras with the same baseline. For this reason, this kind of a camera arrangement has been called by the author of this paper Equal Baseline Multiple Camera Set (EBMCS).

Other sets considered in the research are subsets of the set with five cameras. A set containing four cameras includes all cameras apart from the down one. Similarly, a set with three cameras consists of the central camera, the right and the up one. A set with two cameras is a single stereo camera containing the central camera and the right one.

### 3.2. Calibration and rectification of EBMCS

The set of cameras used in the experiments presented in this paper has been calibrated and rectified with the use of the OpenCV library (http://opencv.org/). Images from each considered stereo camera in EBMCS were processed as if they were images form a stereo camera consisting of a right and a left camera when a right camera is a reference one. Without image transformations, this applies only to stereo camera consisting of the camera 0 and the camera 1. In the research presented in this paper, images from other stereo cameras were transformed in order to match this requirement.

Images from stereo camera 2 were rotated 90 degrees clockwise. Images from stereo camera 3 were flipped making them mirror reflections of original images. Images from stereo camera 4 were rotated 90 degrees counter-clockwise. No transformations were made on images from camera 1. With the use of these transformations, images from every considered stereo camera were calibrated with respect to the same image from a reference, central camera. Therefore, it is possible to use every calibrated pair of images in order to obtain a disparity map that corresponds to points from the image taken by the central camera.

The calibration and the rectification were performed on a series of images of a chessboard with 10 corners in the horizontal dimension and 7 corners in the vertical one. Fig. 2 presents the image of the chessboard. The chessboard was printed on an A4 paper. The size of squares was 24 mm × 24 mm. The series of chessboard images consisted of 10 sets with five images from different

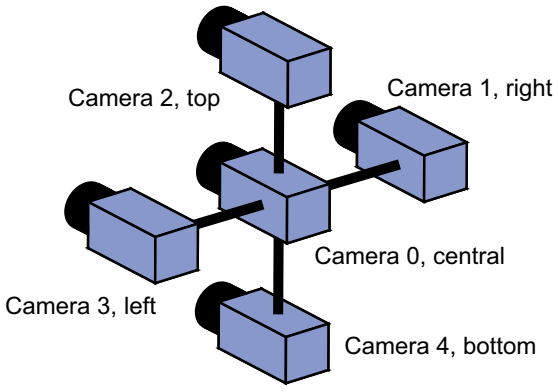


**Fig. 1.** The arrangement of five cameras in Equal Baseline Multiple Camera Set (EBMCS).

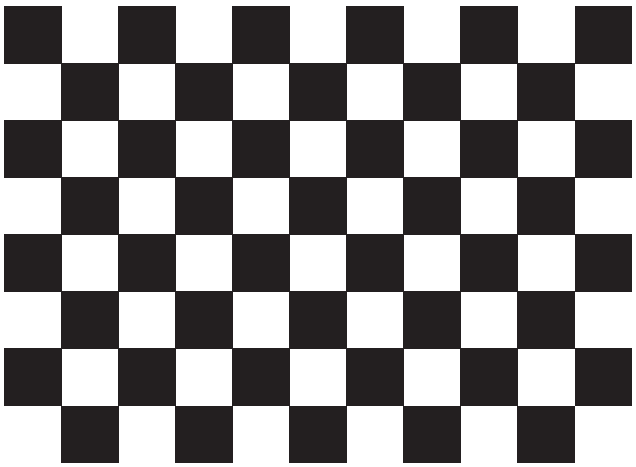

**Fig. 2.** The chessboard used in calibration.

cameras in each set. It is required from calibration images to contain the entire view of a chessboard. The chessboard was placed in 10 different positions and in every position it was visible from all five cameras in the set. Images for calibration were made simultaneously for all cameras.

The calibration parameters were calculated on the basis of the OpenCV implementation of the Zhang's algorithm (Zhang, 2000). Cameras were rectified by the Hartley's algorithm implemented in this library (Hartley, 1999). The rectification included a multicamera adjustment of all cameras in the set. However, in a real device consisting of five cameras there are always some inequalities in distances between each side camera and a central camera. They were also corrected in the process of the rectification.

Apart from the geometric calibration a color calibration was performed on images from EBMCS. In the experiments with this set, images in greyscale were used for making disparity maps. Therefore, the color calibration was in fact the calibration of points' intensities. Intensities were adjusted with respect to characteristic points occurring in images. Intensities of points in images have been modified with the use of triangular filter presented in Eq. (4).

$$\hat{c} = c + \left(1 - \frac{|M - c|}{M}\right)L; \tag{4}$$

where $c$ is an intensity before the calibration, $\hat{c}$ is the intensity after the calibration, $L$ is an intensity modification factor selected with respect to characteristic points in images and $M$ is the middle value in a grayscale range. Typically, the grayscale ranges from 0 to 256, $M$ is than equal to 128.

Image transformations also apply to disparity maps obtained on the basis of input images from every stereo camera taken into account in EBMCS. Disparity maps are obtained from images modified by transformations such as a calibration, a rectification, a rotation and a mirror reflection. As a consequence points of each disparity map correspond to points of a central image after these transformations. However, transformations parameters are different in different stereo cameras. Therefore, the central image is modified variously depending on the stereo camera in which it is used. The research presented in this paper is concerned with merging disparity maps in order to acquire a higher quality map. This requires that maps are unified by making them refer to the same image. The unification of disparity maps is obtained by performing on them transformations reverse to those that were performed on images from which these maps were acquired. Points in all resulting disparity maps correspond to points of the input central image before calibrations and rectifications.

## 4. Merging methods

EBMCS provides a disparity map for each pair of cameras included in the set. These constituent maps are merged into a single disparity map which is the result of using EBMCS. This paper introduces and discusses two methods of merging maps obtained with the use of EBMCS. The first method is an Arithmetic Mean Merging Method (AMMM). The second method is Exceptions Excluding Merging Method (EEMM).

In both methods, the disparity of a point located at some coordinates in the resulting disparity map depends on disparities of points located at the same coordinates in constituent disparity maps. Disparity maps may contain occluded areas which does not have disparities. Therefore, the number of merged disparities in some point of the resulting disparity map may be lower than the number of cameras included in EBMCS. If $N$ is the number of cameras in EBMCS and $M_x$ denotes the number of constituent maps which contain values of disparities in points located at coordinates $\mathbf{x}$, then $M_x \leqslant N$.

In case of the AMMM merging method the disparity of a point $p$ located at coordinates $\mathbf{x}$ in the resulting map is equal to the arithmetic mean of disparities of corresponding points in constituent maps. The mean includes only these maps which contain disparity values at considered points. The formula for calculating disparities in the AMMM merging method is presented in Eq. (5).

$$D_f(\mathbf{x}) = \frac{\sum_{1 \leqslant i \leqslant M_x} D_i(\mathbf{x})}{M_x} \tag{5}$$

where $\mathbf{x}$ denotes coordinates of the considered point, $D_f$ is the resulting value of the disparity, $M_x$ is the number of constituent maps containing disparities at coordinates $\mathbf{x}$ and $D_i$ in the value of the disparity in the constituent map with the index $i$.

It is possible that there will be significant differences between values of disparities located at the same coordinates in different constituent maps. The AMMM method does not exclude any values. However, these differences indicate that at least one constituent map contains an incorrect value of a disparity. In order to eliminate potentially incorrectly disparities, the author of this paper developed the EEMM merging method.

The value of a disparity after performing the EEMM merge is denoted by $F(\mathbf{x})$ where $\mathbf{x}$ are the coordinates of a point $p$ for which disparity is calculated. $F(\mathbf{x})$ depends on each disparity $D_i(\mathbf{x})$ at coordinates $\mathbf{x}$ in a constituent map $i$. In case a map does not contain the disparity in the points the value of $F(\mathbf{x})$ is equal to 0. Function $F(\mathbf{x})$ is calculated differently according to the number of constituent maps containing disparities for the point $p$.

If there is only one constituent map indexed $i$ containing the disparity $D_i(\mathbf{x})$, the value of $F(\mathbf{x})$ is equal to $D_i(\mathbf{x})$. When the number of constituent maps having the disparity at coordinates $\mathbf{x}$ is equal to two, the EEMM merging method calculates the difference between these disparities. The difference is equal to $|D_i(\mathbf{x}) - D_j(\mathbf{x})|$ where $i$ and $j$ are indexes of considered constituent maps.

The merging method specifies a maximum acceptable difference denoted as $Q$. The difference greater than $Q$ shows that the value of the disparity is uncertain. Therefore, the merging method states that the disparity is undetermined and $F(\mathbf{x}) = 0$. If the difference between disparities is not greater than $Q$ then $F(\mathbf{x})$ is equal to the arithmetic mean of disparities $D_i(\mathbf{x})$ and $D_j(\mathbf{x})$ (Eq. (6)).

$$F(\mathbf{x}) = \begin{cases} \frac{D_i(\mathbf{x}) + D_j(\mathbf{x})}{2} & \text{if } |D_i(\mathbf{x}) - D_j(\mathbf{x})| \leqslant Q \\ 0 & \text{if } |D_i(\mathbf{x}) - D_j(\mathbf{x})| > Q \end{cases} \tag{6}$$

In case of merging three disparities $D_i(\mathbf{x}), D_j(\mathbf{x})$ and $D_k(\mathbf{x})$ from different constituent maps differences are calculated between each two disparities. There are four versions of calculating $F(\mathbf{x})$ according to results of comparing differences with the parameter $B$ such

that $B = Q/2$ (Eq. (7)). Considering that three measurements are available the condition of maximum acceptable difference is set more strictly than in case of having only two disparities. When all three differences are not greater than $B$ then the resulting disparity for the final disparity map is equal to the arithmetic mean of all disparities. When there is one difference greater than $B$, then $F(\mathbf{x})$ is equal to the disparity for which both other conditions of maximum acceptable difference are fulfilled. In case of two differences greater than $B$ the result is equal to the arithmetic mean of disparities for which the difference is not greater than $B$. The last case occurs when all three differences are greater than $B$. In this case the resulting value of disparity is undetermined.

$$F(x) = \begin{cases} \frac{\sum_{v \in i,j,k} D_v(\mathbf{x})}{3} & \text{if } \begin{array}{l} |D_i(\mathbf{x}) - D_j(\mathbf{x})| \leqslant B \\ |D_i(\mathbf{x}) - D_k(\mathbf{x})| \leqslant B \\ |D_j(\mathbf{x}) - D_k(\mathbf{x})| \leqslant B \end{array} \\ D_i(\mathbf{x}) & \text{if } \begin{array}{l} |D_i(\mathbf{x}) - D_j(\mathbf{x})| \leqslant B \\ |D_i(\mathbf{x}) - D_k(\mathbf{x})| \leqslant B \\ |D_j(\mathbf{x}) - D_k(\mathbf{x})| > B \end{array} \\ \frac{D_i(\mathbf{x}) + D_j(\mathbf{x})}{2} & \text{if } \begin{array}{l} |D_i(\mathbf{x}) - D_j(\mathbf{x})| \leqslant B \\ |D_i(\mathbf{x}) - D_k(\mathbf{x})| > B \\ |D_j(\mathbf{x}) - D_k(\mathbf{x})| > B \end{array} \\ 0 & \text{if } \begin{array}{l} |D_i(\mathbf{x}) - D_j(\mathbf{x})| > B \\ |D_i(\mathbf{x}) - D_k(\mathbf{x})| > B \\ |D_j(\mathbf{x}) - D_k(\mathbf{x})| > B \end{array} \end{cases} \tag{7}$$

The last case in the EEMM merging method occurs when there are four constituent maps $i, j, k, l$ containing disparities in a point at coordinates $\mathbf{x}$. In this situation, the merging method first sorts values of disparities from different maps and afterward it excludes the extreme values. The arithmetic mean is then calculated from two remaining disparities. This mean is the result of the merging method. The formula for merging four disparities is presented in Eq. (8).

$$F(\mathbf{x}) = \frac{D_j(\mathbf{x}) + D_k(\mathbf{x})}{2} \text{ if } \begin{array}{l} D_i(\mathbf{x}) \leqslant D_j(\mathbf{x}) \leqslant \\ D_k(\mathbf{x}) \leqslant D_l(\mathbf{x}) \end{array} \tag{8}$$

## 5. Testbed

The author of this paper prepared three data sets for the purpose of making experiments with the designed merging methods. Each set consists of five images taken from different points of view and ground truth.

### 5.1. Data sets

Images used in tests were made with the set of cameras arranged in the EBMCS configuration described in Section 3.1. The imaging devices were web cameras model MS LifeCam Studio with the 1080p HD sensor. Cameras were fixed to an aluminum construction mounted on a retort stand. The image of the real EBMCS is presented in Fig. 3.

The distance between each side camera and the central camera was equal to 50 mm. The producer of cameras does not provide information on the focal length of lens in this model. However, the focal length in pixel units and other data needed to convert disparity maps to depth maps can be acquired on the basis of calibration data. The OpenCV library estimated that the focal length is equal to 972 in pixel units. This value was set in the camera matrix calculated by the library on the basis of calibration images described in Section 3.2 (Bradski and Kaehler, 2008).

**Fig. 3.** Real EBMCS used in the experiments.

All images were made in an out-door environment under natural light conditions. The author of this paper prepared for the experiments six data sets containing images of different plants including strawberries, redcurrants and cherries. Two data sets were prepared for each kind of a plant. Strawberry data sets are marked in this paper with $ST_1$ and $ST_2$. Redcurrant sets are marked with $RC_1$ and $RC_w$. Sets with images of cherries are marked with $CH_1$ and $CH_2$. All images contained views of ripe fruits on plants as presented in Figs. 4–6.

Ground truth was prepared for each set of images. Points of ground truth correspond to points of a central image. Ground truth covers a middle part of a central image containing objects which are also included in all other images from EBMCS with five cameras. When two images from a stereo camera are used, parts of each image located near one of its sides contain views of objects which are not visible in the other image. Stereo matching algorithms make disparity maps for objects and their parts which are visible in both input images. In case of using EBMCS with five cameras this issue applies to every considered pairs of images. In the vicinity of every edge of a central image there is an area which is not visible in at least one of side images. Therefore, a central image consists of a matching area in the middle part of the image and a margin around this area. The size of ground truth is the same as the size of a matching area.

The sizes of matching areas, margin sizes and ranges of disparities occurring in ground truth are presented in Table 1. Experiments presented in this paper were performed for all points located within the matching area in every data set. Therefore, the total number of points used in the experiments is equal to 172400.

The size of input images presented in Figs. 4–6 results from the size of matching areas and the side of margins. For example, in case of the first set of strawberry plant $ST_1$, the matching area is a rectangle of size $240 \times 180$ points. The margin is 100 points wide. Therefore, images have size of $440 \times 380$ points. Ground truth for this set contains disparities in the range from 1 to 28.

There are four factors which affects the minimal required size of the margin:

- the range of real disparities in the image set
- ranges of disparities checked by stereo matching algorithms
- sizes of aggregating windows used in stereo matching algorithms
- other configurations of stereo matching algorithms

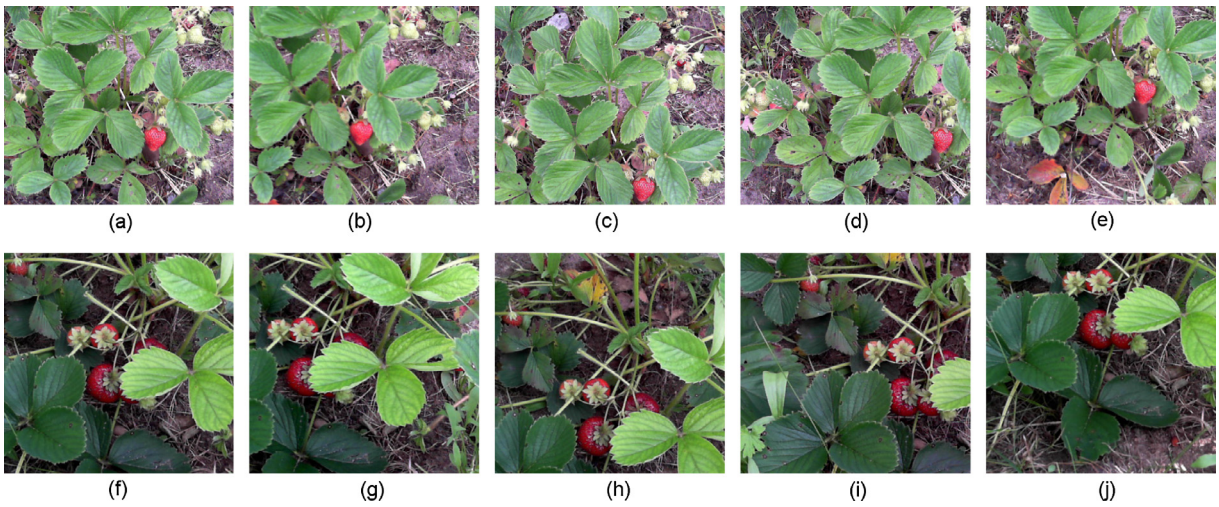


**Fig. 4.** Images of strawberry data sets $ST_1$ ((a)-(e)) and $ST_2$ ((f)-(j)); (a, f) – central; (b, g) – right; (c, f) – up; (d, i) – left; (e, j) – down.

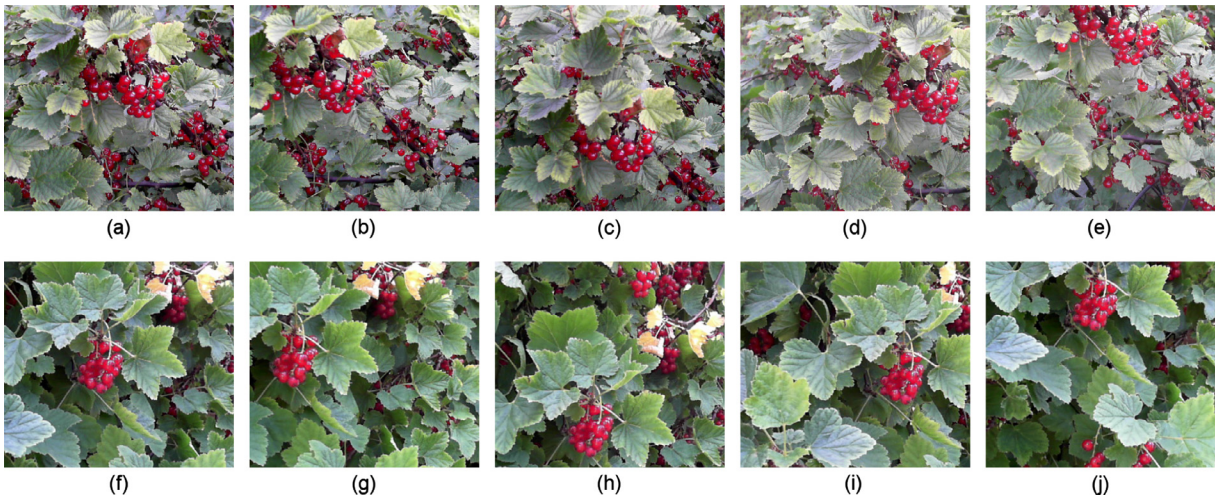


**Fig. 5.** Images of redcurrant data sets $RC_1$ ((a)-(e)) and $RC_2$ ((f)-(j)); (a, f) – central; (b, g) – right; (c, f) – up; (d, i) – left; (e, j) – down.

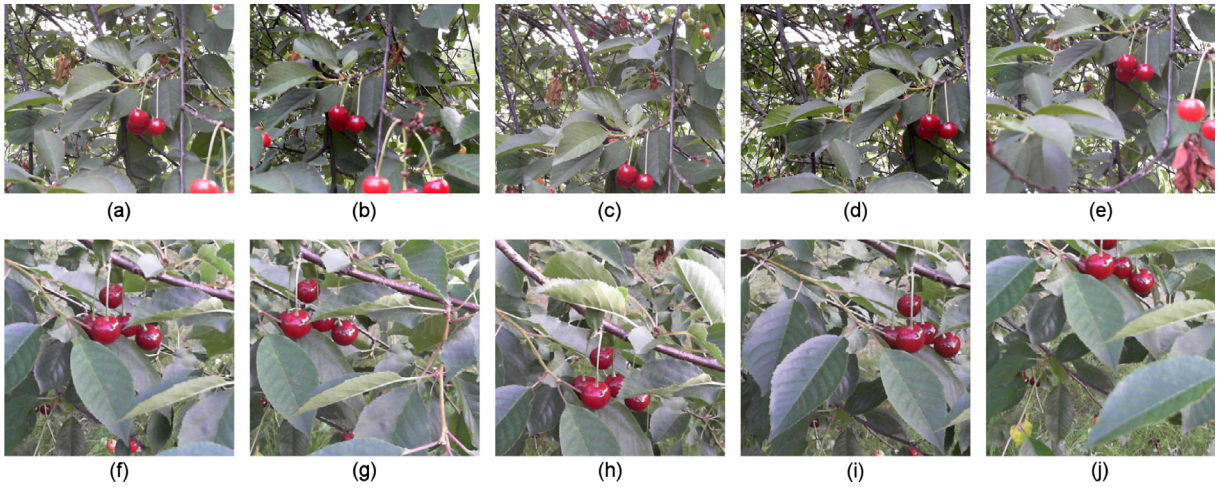


**Fig. 6.** Images of cherry data sets $CH_1$ ((a)-(e)) and $CH_2$ ((f)-(j)); (a, f) – central; (b, g) – right; (c, f) – up, (d, i) – left; (e, j) – down.

The size of the margin needs to be wider than the minimum margin size which makes it possible to acquire disparities. If the margin size is lower than the maximum real disparity, then it is possible that a side image does not contain the view of an object visible within the matching area of the reference image. This problem occurs when a points with maximum disparity is located near

**Table 1**
Parameters of test data sets.

| Set ID | Matching size | Margin | Disparity range |
|---|---|---|---|
| $ST_1$ | 240 × 180 | 100 | 1–28 |
| $ST_2$ | 140 × 125 | 110 | 19–45 |
| $RC_1$ | 220 × 170 | 130 | 31–79 |
| $RC_2$ | 120 × 95 | 110 | 11–35 |
| $CH_1$ | 270 × 180 | 100 | 40–69 |
| $CH_2$ | 130 × 110 | 110 | 27–58 |

the edge of the reference image. In this case a side image does not contain a point corresponding to a matched point because the location of the corresponding point is beyond borders of the side image. Moreover, stereo matching algorithms consider aggregating windows which contains vicinities of points. The side image, apart from a corresponding point needs to contain points surrounding it. The size of an aggregating window depends on the stereo matching algorithm and its configuration. The minimum margin size needs to be not lower than the range of checked disparities enlarged by the size of an aggregating window.

The minimum margin size also depends on the range of disparities analyzed by stereo matching algorithms. This range is set to be wider than the range of real disparities. The size of the margin needs to make it possible for the algorithm to verify disparities within the analyzed range. Moreover, stereo matching algorithms consider the content of the entire image in the matching process. For example, ELAS uses the whole image to select support points defining the range of checked disparities in this algorithm. Additionally expanding the margin is beneficial for these calculations. In general, the margin cannot be too narrow because it makes it impossible for the matching algorithm to appropriately generate the disparity map. However, it is advantageous when the size of the margin is greater than the minimum required value. In case of every data set included in Table 1 margins are wide enough to perform experiments presented in this paper.

Test data contain also ground truth. Ground truth was obtained as a result of manual analysis of image sets. Acquiring ground truth manually requires recognizing the same parts of plants in different images. The author of this paper manually matched points of central images with corresponding points in side images. The identified differences in locations of corresponding points were selected as values of disparities included in ground truth. This process of preparing ground truth is very time consuming. However, using this method results in obtaining dense ground truth data. Scharstein et al. described different methods for obtaining ground truth (Scharstein et al., 2014). Ground truth can also be obtained with the use of various kinds of equipment for measuring distances including 3D structured-light scanners (Jang et al., 2013). However there are no devices which can automatically provide a perfect quality of dense ground truth data. Therefore, ground truth is obtained manually or manual corrections are made on data acquired from measuring devices.

Values of disparities in ground truth used in this paper are determined on the basis of images which were calibrated and rectified. These processes transform images and relocate their content. Thus, ground truth made for images before the calibration and the rectification differ from ground truth for calibrated and rectified images. All stereo matching algorithms tested in the research presented in this paper were executed for calibrated and rectified images. Therefore, both the results of these algorithms and ground truth contain values of disparities extracted from the same kind of images. Although values of disparities in ground truth are based on calibrated and rectified images, the locations of these values correspond to points of central images without these transformations. Such ground truth can be applied for estimating the quality of all disparity maps obtained from stereo cameras included in EBMCS.

Ground truth also contains parts called occluded areas in which it is not possible to determine real values of disparities. One of the main reasons is that these areas are visible only in a central image to which ground truth refers. Occluded areas are located in the background of the scene and they are hidden behind objects placed in the foreground from points of view of side cameras. In ground truth presented in this paper, occluded areas are denoted with black color.

Ground truth for images of the strawberry plant $ST_1$, the redcurrant plant $RC_1$ and the cherry tree $CH_1$ are presented in Fig. 7a, f and k, respectively. Intensities of points in these figures were normalized to a full grayscale range in order to make them more visible. Brighter points represent greater values of disparities.

### 5.2. Quality metrics

Szeliski et al. described two metrics for estimating the quality of disparity maps: the root-mean-squared error (RMS) and the percentage of bad matching pixels (BMP) (Scharstein and Szeliski, 2002). Both metrics calculate differences between values in a disparity map and corresponding values in ground truth. RMS extracts the quadratic mean of these differences. The formula of RMS for estimating the quality of disparity maps is presented in Eq. (9).

$$RMS = \sqrt{\frac{1}{N}\sum_{\mathbf{x}} |D_M(\mathbf{x}) - D_T(\mathbf{x})|^2} \tag{9}$$

where $\mathbf{x}$ are coordinates of a point, $D_M(\mathbf{x})$ is the disparity of the point in the disparity map, $D_T(\mathbf{x})$ is the disparity in ground truth and $N$ is the total number of points.

The formula for the percentage of bad matching pixels is given in Eq. (10). This metric calculates differences between disparities in the disparity map and in ground truth and then compares these differences with a border value defining the disparity error tolerance. Points for which the difference is not greater than the border value are considered to be matched correctly. Points are assumed to be matched incorrectly otherwise.

$$BMP = \frac{1}{N}\sum_{\mathbf{x}} (|D_M(\mathbf{x}) - D_T(\mathbf{x})| > Z) \tag{10}$$

where $Z$ is the border value and other symbols are the same as in Eq. (9).

Ground truth contains areas with values of real disparities and occluded areas where disparities are undetermined. The usage of EBMCS has an influence on both of these areas. The author of this paper used different metrics to measure the influence of EBMCS on the matching quality depending on the area type.

In the experiments presented in this paper, a modified version of the BMP metric was used. The modification occurs in using BMP only for points which have disparities in ground truth. In the modified version of BMP, the value $N$ in Eq. (10) is equal to the number of points containing disparities in ground truth instead of total number of point.

Apart from BMP, another metric was used for measuring the stereo matching quality in areas of disparity maps corresponding to occluded areas in ground truth. The metric was called the percentage of bad matching pixels in occluded area (BMO). The formula of the BMO metric is given in Eq. (11).

$$BMO = \frac{1}{N_B}\sum_{\mathbf{x}} (D_M(\mathbf{x}) \neq 0 \wedge D_T(\mathbf{x}) = 0) \tag{11}$$

where $N_B$ is the number of points in the occluded area and other symbols are the same as in Eq. (9).

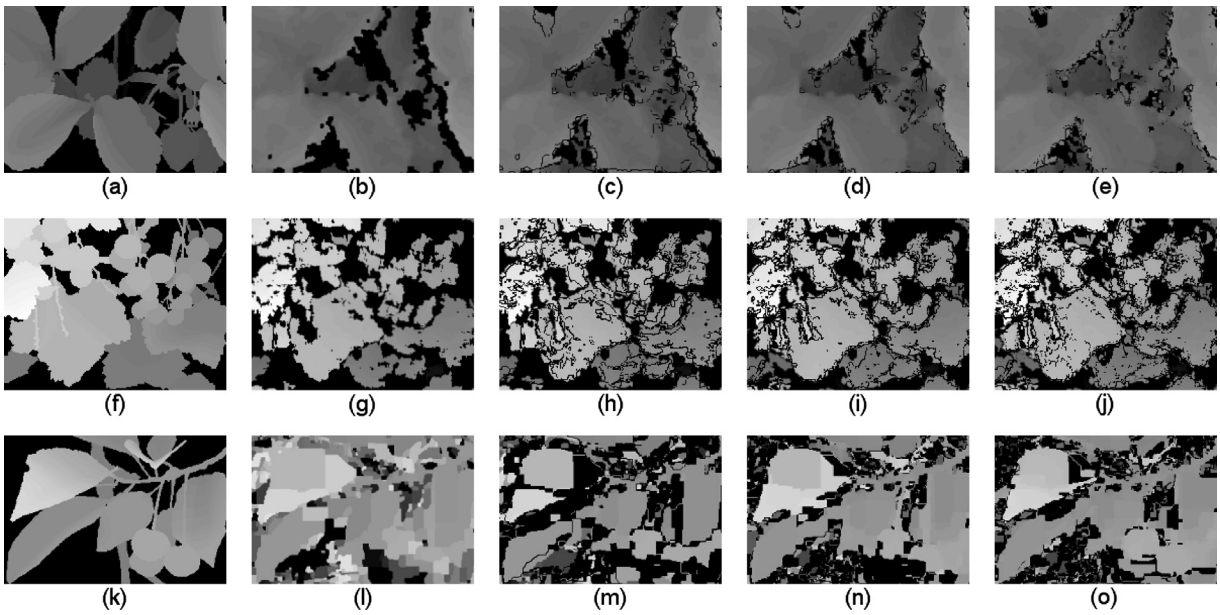


**Fig. 7.** Ground truth and disparity maps for the number of cameras between two and five; the strawberry plant $ST_1$: (a) ground truth, (b)-(e) results for the ELAS algorithm; the redcurrant plant $RC_1$: (f) ground truth, (g)-(j) results for the StereoBM algorithm; the cherry tree $CH_1$: (k) ground truth, (l)-(o) results for the GC Swap algorithm.

A point in a disparity map is matched incorrectly if it contains a value of disparity ($D_M(\mathbf{x}) \neq 0$) but the corresponding point in ground truth is within the occluded area ($D_T(\mathbf{x}) = 0$). The BMO metric takes into account only these points that are covered by the occluded area. The metric is equal to the percentage of bad matching points in the occluded area.

A next metric used for measuring the quality of disparity maps obtained from EBMCS is coverage level (COV). This metric is calculated by dividing the number of points for which a stereo matching algorithm found disparities by the total number of processed points. The coverage level shows the percentage of points which were not classified as an occluded area. The formula for the COV metric is presented in Eq. (12).

$$COV = \frac{N_L}{N} \tag{12}$$

where $N_L$ is the number of points for which disparities were found and $N$ is the total number of points processes by a stereo matching algorithm.

## 6. Experiments

A series of experiments was performed in order to examine AMMM and EEMM merging methods used with EBMCS. Experiments were based on data sets described in Section 5.1. The main purpose of experiments was to present the performance of two merging methods presented in this paper with regard to the number of stereo cameras included in EBMCS. Experiments were also performed to compare the performance of different stereo matching algorithms used with EBMCS. Results of experiments show the optimal combination of three following factors: the number of stereo cameras in EBMCS, the merging method type and the type of the stereo matching algorithm. Additionally, experiments include the selection of the parameter $Q$ used in EEMM (Section 4) and tests of the algorithm proposed by Park and Inoue (Section 2.3).

Experiments considered EBMCS with the number of cameras varying from two to five. Sets with less than five cameras contained a central camera and a subset of side cameras from the five camera set. A subset of stereo cameras consisting of a central camera and sides one can be selected differently from a five camera EBMCS because stereo cameras included in EBMCS are distinguishable. Subsets used in the experiments consisted of cameras selected as described at the end of Section 3.1.

In general, all stereo cameras included in EBMCS acquire disparity maps of the same quality although cameras are distinguishable. It is possible that some side cameras have more advantageous viewpoints in case of some parts of images. It depends on the shape and the structure of viewed plants. However, every stereo camera is in fact the same device rotated by different angles. This theoretical expectation was additionally verified by the author of this paper by comparing disparity maps obtained by using each single stereo camera considered in EBMCS. These results show that there is no stereo camera which always provides a disparity map with the highest quality.

All algorithms except from GC Swap and GC Expansion always created the same disparity maps for the same input data. However, in case of GC Swap and GC Expansion there were differences between subsequent executions of these algorithms in the same configurations. Changes were caused by the design of these algorithms which includes random selections in data processing order (Boykov et al., 2001). Differences also occurred in the results of quality metrics used in assessing qualities of these maps. Therefore, in order to obtain typical results these algorithms were executed 10 times for the same input data and the same stereo matching parameters. The results of quality metrics presented in this paper for GC Swap and GC Expansion are average values obtained from these executions.

### 6.1. Application of stereo matching algorithms

Experiments were performed with 8 stereo matching algorithms described in Section 2.2: ELAS, StereoBM, StereoSGBM, StereoVar, GC Swap, GC Expansion, BP-M and TRW-S. Results of stereo matching algorithms depend on their input parameters. Modifying input parameters can either improve or deteriorate results with regard to a type of images processed by these algorithms. Images differ in features such as contrast, brightness and

textures of visible objects. This kind of differences also occurs in images of plants. The quality of a disparity map can be improved by adjusting parameters to specifics of a particular set of images. However, an algorithm intended for use with different images needs to provide high quality disparity maps regardless of characteristic features of input data. Thus, designers and developers of stereo matching algorithms select parameters which are suitable for all types of images from stereo cameras. This kind of default values were used in the experiments presented in this paper. Ranges of disparities considered in calculations were the only input parameters changed from default.

In case of GC Swap, GC Expansion, BP-M and TRW-S, default parameters were used from the implementation provided by Middlebury Stereo Vision Page (Section 2.2). The implementation takes the following arguments: use Birchfield/Tomasi costs, use squared differences, truncated differences, the smoothness exponent, the maximum value of smoothness term, the weight of smoothness term, the intensity gradient cue threshold, the smoothness multiplication factor and the scale factor (Boykov et al., 2001; Tappen and Freeman, 2003; Wainwright et al., 2005; Kolmogorov, 2006). In case of StereoBM, StereoSGBM and StereoVar parameters were acquired from the exemplary code provided with the OpenCV library. StereoBM and StereoSGBM share common attributes such as the truncation value, the window size, the uniqueness ratio, the speckle window size, the speckle range and the maximum allowed difference (Bradski and Kaehler, 2008). Moreover, StereoBM has the texture threshold parameter. StereoSGBM has also parameters P1 and P2 for controlling the disparity smoothness. The OpenCV implementation of the StereoVar algorithm accepts the following arguments: the number of layers, the image scale, the number of iterations, the size of the pixel neighborhood, the standard deviation of the Gaussian, the smoothness parameter, the threshold parameter for edge-preserving smoothness, variants for smoothness, the type of the multigrid cycle, use the input flow as the initial flow approximation, use the histogram equalization, use the smart iteration distribution and use the median filer (Bradski and Kaehler, 2008). The ELAS algorithm is delivered with two versions of its configuration (Geiger et al., 2011). The first configuration is the default setting for a robotics environment and the second one is intended for use with Middlebury test data sets (Section 2.2). In the experiments presented in this paper the first configuration was used.

### 6.2. Range of disparities

Apart from using default values it was necessary to set the range of disparities checked by algorithms. The ELAS algorithm does not require a disparity range as input because the algorithm estimates the range by itself. StereoBM, StereoSGBM and StereoVar algorithms require setting the disparity range boundaries as multiplications of the value 16. The maximum value of disparity was equal to 48 for $ST_1$ and $RC_2$ sets. It was equal to 64 for $ST_2$ and $CH_2$ sets and it was equal to 80 for $RC_1$ and $CH_1$ sets. In algorithms GC Swap, GC Expansion, BP-M, TRW-S and in the algorithm proposed by Park and Inoue it is not necessary that boundaries are divisible by 16. The upper disparity boundary was set to the following values: 40 for $ST_1$, 50 for $RC_2$, 60 for $ST_2$, 70 for $CH_1$, 70 for $CH_2$ and 80 for $RC_1$.

The range of disparities checked by a stereo matching algorithm needs to be wider than the range of real disparities occurring in images in order to make it possible for a stereo matching algorithm to match corresponding areas in images from different cameras. However, the wider is the verified disparity range, the harder it is for a stereo matching algorithm to correctly acquire disparities. The range of disparities can be estimated by detecting characteristic points in all input images. Locations of these points indicate the disparity range. This method is used in the ELAS algorithm (Geiger et al., 2011). The range of disparities can also be assessed on the basis of approximate distance between cameras and viewed objects. In some applications the approximate distance is determined by conditions in which cameras are used. For example, if cameras are mounted on a fruit harvesting robot, then the distance can be determined with respect to the size of a robot and the location of cameras.

In the experiments presented in this paper the disparity range was selected with regard to the range of disparities occurring in ground truth presented in Table 1. In case of every data set the upper boundary was set to a value which was greater than the maximum disparity present in ground truth. The upper boundary was greater than the maximum disparity in different extent for various data sets. Selecting different extents reflects conditions in field applications in which there can be different ranges of possible distances between a camera set and plants.

### 6.3. Parameters Q and Z

The EEMM merging method proposed in this paper requires setting the acceptable difference rate $Q$ (Section 4). This parameter was set to 9. For this value two merged disparities are within the margin of error if they differ no more than $Q$. In case of three input disparities EEMM is selecting maps for which disparities differ not more than a half of the $Q$ parameter, i.e. 4.5.

The level of the $Q$ parameter was selected experimentally. The author conducted experiments that involved measuring average values of BMP, BMO and COV metrics with regard to the value of the parameter $Q$. Average values were calculated from results of all considered stereo matching algorithms. Values of $Q$ were verified separately for three EBMCS configurations consisting of three, four and five cameras. Results showed that there was a major decrease in disparity maps quality when the parameter $Q$ was less than 6. For these values the BMP metric was high, e.g. for 5 cameras and $Q = 1$ BMP was equal to 58.54%. The COV metric in this configuration was equal to 44.73% and BMO was 24.7%. When $Q$ was greater than 6 differences in average values of quality metrics were not as significant as in case of differences for $Q$ lower than 6. BMP calculated for 5 cameras was in the range between 19.7% and 17.57% when $Q$ was between 7 and 16. BMP was dropping with the increase of $Q$. The COV metric for this range was between 88.56% and 93.65%. BMO was between 52.78% and 66.7%. Both COV and BMO were greater for greater values of $Q$. Results for configurations with four and three cameras showed a similar characteristic. In general, the rise in $Q$ value caused a minor decrease in the BMP metric and it caused an increase in COV and BMO metrics. Therefore, the parameter $Q$ was selected to be beyond values for which the BMP metric indicated low quality of maps, however the parameter was set to be low in order not to increase the BMO metric. $Q$ equal to 9 matched both these conditions. This value was used in experiments.

Another parameter which needs to be set in the experiments is the border value $Z$ in the percentage of bad matching pixels metric (Section 5.2). $Z$ is a quality metric parameter for comparing results of different stereo matching algorithms. If the parameter is more rigorous, i.e. its value is lower, then the metric will indicate more errors in evaluated disparity maps. If it is higher, then results will be regarded as more correct. It is essential to compare all disparity maps with the use of the same metrics and their parameters. The author used the value of $Z$ equal to 2, however the comparison can also be performed for other values of this parameter.

### 6.4. Results of experiments

The main purpose of experiments is to examine the influence of the number of cameras included in EBMCS on the quality of disparity maps obtained with the use of EEMM and AMMM disparity maps merging methods. Fig. 8a-f presents results of the experiments for both of these methods. Horizontal axes correspond to numbers of cameras in EBMCS and vertical axes show values of quality metrics. Charts present results for 8 stereo matching algorithms including ELAS, StereoBM, StereoSGBM, StereoVar, GC Swap, GC Expansion, BP-M and TRW-S. Each point referring in charts to these algorithms represents the average value of results obtained from all data sets. The experiment was performed on all six data sets described in Section 5.1. Charts also include lines marked with *Average* which show average values calculated on the basis of all stereo matching algorithms considered in the experiments.

Fig. 8a-c presents results for the AMMM merging method. The results demonstrated that using this method does not lead to the improvement in the quality of disparity maps. The increase in the number of cameras does not implicate the decrease in the value of the BMP metric as shown in Fig. 8a. The greatest improvement in the value of this metric occurred in case of the ELAS matching algorithm. The value was 24.67% when two cameras were used and it was reduced to 19.74% for the configuration with three cameras. However, in case of the StereoSGBM matching algorithm, results for four and five cameras were even worth than results for two cameras. In general, results have not changed significantly in case of most of matching algorithms.

Values of the coverage metric improved or remain constant with the increase in the number of cameras for all algorithms. The greatest improvement occurred for algorithms ELAS, StereoBM and StereoSGBM. However, the improvement was related to deterioration of matching quality in areas classified as occluded. Using a greater number of cameras caused that values of the percentage of bad matching pixels in occluded areas increased for all these stereo matching algorithms.

Algorithms StereoVar, GC Swap, GC Expansion, BP-M and TRW-S generate results for almost every point of a disparity map. Values of the COV metrics for these algorithms are above 99.33% if two cameras are used. Therefore, these algorithms identify occluded areas in a low extent. Using more cameras with the AMMM method does not cause occurrence of occluded areas. If five cameras are used, no area is classified as occluded resulting in the COV metric equal to 100%. As a consequence, the error rate in occluded areas measured with the BMO metric is also equal to 100%. Therefore, AMMM is not suitable for identifying occluded areas.

Better results than in case of AMMM were obtained for the matching algorithm proposed by Park and Inoue (Section 2.3). The algorithm does not use a merging method such as AMMM or EEMM. The algorithm was designed for five cameras and it does not have a version for a different number of cameras. In the experiments average values of quality metrics were calculated on the basis of the results obtained for all six data sets. The BMP metric was equal to 18.38%. COV and BMP metrics were both equal to 100%. These results are better than in case of using the TRWS algorithm which provided the best results for the AMMM merging method. However, results of the algorithm proposed by Park and Inoue were worse than results of the EEMM merging method.

In general, the EEMM method makes it possible to reduce the percentage of bad matching pixels metric by increasing the

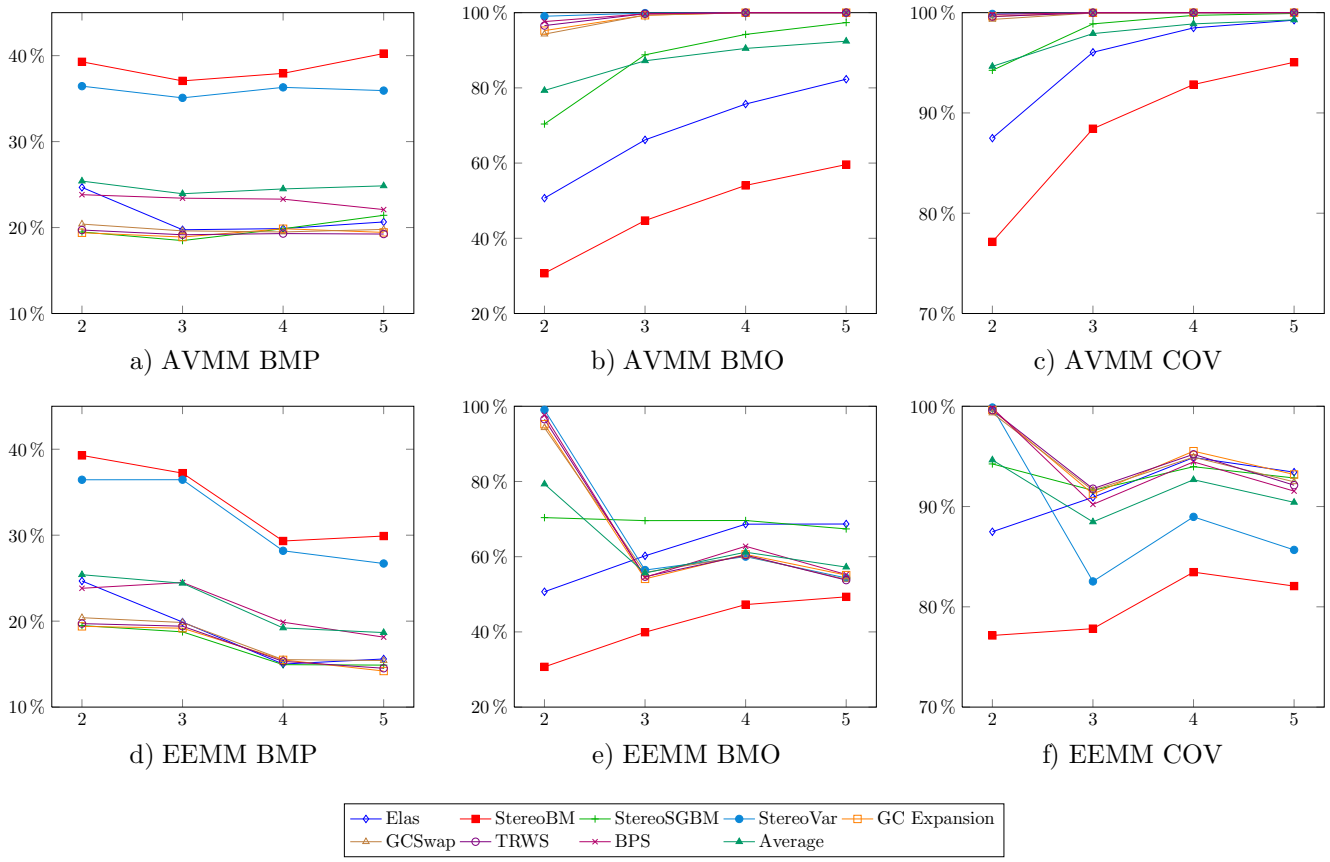


**Fig. 8.** Results of Arithmetic Mean Merging Method (a-c) and Exceptions Excluding Merging Method (d-f).

number of cameras. Fig. 8d presents the results. The value of BMP was worse for disparity maps based on three cameras than for maps obtained from two cameras in case of some matching algorithms. However, taking advantage of four or five cameras always produced results with better BMP than using two cameras.

The greatest improvement in BMP occurred for the ELAS algorithm. BMP was equal to 24.67% for two cameras and it was equal to 15.01% when four cameras were used. Therefore, BMP reduced 39.17% of its value for two cameras. The lowest improvement occurred in case of the StereoSGBM algorithm. BMP reduced from 19.48% to 14.89% when five cameras were used instead of two ones. For this algorithm, BMP reduced 23.6% of its value for two cameras. On average, BMP for five cameras improved by 26.55% of its average value for two cameras with the use of the EEMM method. The best results were obtained with the GC Expansion algorithm used with five cameras. BMP was then equal to 14.2%.

Fig. 7b-e, g-j, l-o presents three series of disparity maps obtained with the use of EEMM. Each series consists of maps made with the use of subsequent number of cameras in the range from two to five. The first series shows maps obtained for the strawberry $ST_1$ set with the use of the ELAS matching algorithm. The second series shows maps of the redcurrant plant $RC_1$ obtained with the use of the StereoBM algorithm. The last series shows disparity maps for $CH_1$ made by the GC Swap algorithm. Intensities were normalized in order to make maps more visible. Intensities span along a full grayscale range similarly to images of ground truth (Section 5.1).

The area of a disparity map is the same as the matching area for each data set. Matching areas are located in middle parts of central images presented in Figs. 4–6. Central images consist of matching areas and margins as described in Section 5.1. Sizes of matching areas and margins are presented in Table 1. Although disparity maps presented in Fig. 7 are in grayscale, they show disparities for both leaves of plants and fruits. Brighter points mark parts of plants located closer to the viewpoint. Matching areas in all considered data sets contain views of leaves and at least one fruit. The matching area of the strawberry data set $ST_1$ (Fig. 4a-e) contains images of a single ripe strawberry. In disparity maps 7b-e it is visible in right bottom corners of maps. In case of the redcurrant plant $RC_1$ (Fig. 5a-e) a large majority of fruits are visible in top right parts of disparity maps 7g-j. There are also some fruits in top left parts. The matching area of the cherry data set $CH_1$ (Fig. 6a-e) shows a bunch of three cherries with one of them partly hidden. Disparities for these fruits are present in right bottom parts of disparity maps presented in Fig. 7l-o.

Fig. 8e shows the influence of EEMM on the percentage of bad matching pixels in occluded areas. Fig. 8f presents coverage levels for the EEMM method. The greatest differences in values of these metrics occur in cases of stereo matching algorithms which identify occluded areas in a low extent (StereoVar, GC Swap, GC Expansion, BP-M and TRW-S). These algorithms have the coverage metric above 99.33% if they are used with a single stereo camera. However, the coverage level drops to below 91.78% when they are used with three cameras and EEMM. The merging method causes an extraction of occluded areas. Algorithms which do not appropriately discover occluded areas are able to find these areas with the use of EEMM. This is a major advantage of using EEMM because the method provides information that disparity value is unavailable instead of providing false data.

The drop in the coverage level corresponds to the drop in the BMO metrics. This metric is significantly reduced by over 34.61% because of recognizing occluded areas. There are some fluctuations in values of the COV metric and the BMO metric for the number of cameras between three and five. However, the value of the coverage is between 77.82% and 95.52% for all tested algorithms when the number of cameras is greater than two. The BMO metric is between 39.92% and 69.62% in these configurations.

EEMM used with multiple cameras has also some influence on algorithms that identify occluded areas in a large extent with the use of two cameras. Using more cameras causes the increase in the coverage level for ELAS and StereoBM. In case of StereoBM the coverage rises from 77.15% to 82.07% when five cameras are used. For ELAS it rises from 87.49% to 93.43% with the use of five cameras. However, the value of the BMO metric is also higher if more than two cameras are used. Increasing the number of cameras for these algorithms causes that more points in occluded areas are matched incorrectly. The increase is higher than 17.97% when five cameras are used instead of two ones. EEMM has the lowest influence on values of BMO and COV metrics for the StereoSGBM algorithm. In general, the influence of EEMM on algorithms ELAS, StereoBM and StereoSGBM is not as significant as in case of algorithms that do not identify occluded areas in a large extent. The most valuable advantage of the EEMM method is making it possible to reduce the value of the BMP metric for all tested algorithms and to identify occluded areas in algorithms that mainly do not discover these areas.

## 7. Agricultural applications

EBMCS is designed for robotic fruit harvesting. EBMCS has a function of a module responsible for estimating distances similarly as stereo cameras, TOF cameras and structured-light 3D scanners (Section 2.1). Robotic harvesters use this kind of devices for estimating distances between them and fruits ready for picking. The distance estimating equipment can be placed on different parts of a robot. These devices are placed either on a robotic arm for picking up fruits or on the main vehicle of a robot. Possible locations of EBMCS on a harvesting robot are presented in Fig. 9.

Placing the module on the robotic arm provides the robot with detailed data on distances between fruits localized on plants and the robotic fruit gripper. When the module is placed on the vehicle,

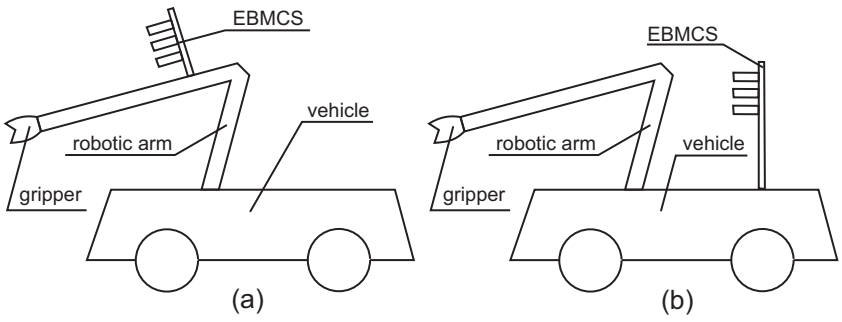


**Fig. 9.** Possible location of EBMCS on a harvesting robot: (a) mounted on a robotic arm, (b) mounted on a main vehicle.

the information on distances is less detailed however the viewing area is wider. EBMCS can be mounted in both of these locations.

EBMCS is exposed to a variety of noises when it is used with a harvesting robot operating in the field. Noises can be caused by vibrations of a robot, in particular when it is equipped with a combustion engine. Foremost, the construction of EBMCS needs to be stable enough to ensure that mutual locations of cameras included in EBMCS remain unchanged despite any vibrations. The other problem with vibrations is such that they affect the process of taking images. If vibrations are high, images can be blurred. Nevertheless, a camera working in daylight uses a fast shutter speed and therefore vibrations do not significantly influence the quality of images. In case of using EBMCS with low intensity of natural light an additional light source is necessary.

A greater problem than vibrations are weather conditions. Strong wind severely moving plants can lead to similar problems as vibrations. It may cause the necessity for using a fast shutter in order take sharp images of moving objects. However, even minor wind is problematic because it sets leaves and other parts of plants in constant motion. If cameras included in EBMCS take images in different moments of time, they will capture parts of plants in different positions. It generates errors in disparity maps obtained from these images. Therefore, it is crucial that all cameras take images simultaneously. If cameras are controlled by a single computer it is enough to simultaneously trigger all cameras. Images can be afterward successively transferred from cameras to the computer.

Another weather condition affecting a field installation of EBMCS is the rain. If a harvesting robot with EBMCS is expected to be operating in rain, it needs to be fully waterproof. Images taken in the rain depict plants less precisely than images made in clear weather. Consequently, it will have a negative effect on the quality of disparity maps obtained from EBMCS. The rain does not exclude the possibility of using EBMCS however EBMCS is in general intended for use when the rain is not falling.

## 8. Conclusions

Applying the research presented in this paper to robotic harvesters will increase the precision of locating fruits for harvest. The main contribution of the research presented in this paper is the observation that Equal Baseline Multiple Camera Set used with Exceptions Excluding Merging Method improves results of stereo matching algorithms used for purposes of the distance estimation. In case of the percentage of bad matching pixels metric the improvement varies among eight tested algorithms from 23.6% to 36.7% in comparison to results obtained with the use of a single stereo camera. On average, it is 26.55%. The best quality of the distance estimation was obtained for the GC Expansion algorithm used with EEMM and five cameras.

Moreover, EEMM identifies areas in which it is not possible to determine the distance. Some stereo matching algorithms provide false data about the distance estimation instead of information that the distance is undeterminable. When these algorithms are used with EEMM this information is available. This feature appears for example in case of the GC Expansion algorithm.

In further work, we plan to experiment with more kinds of plants and we plan to develop other methods of improving the quality of disparity maps by taking advantage of EBMCS. In particular, we plan to investigate merging disparity maps obtained with the use of different stereo matching algorithms. Each stereo camera consisting on EBMCS can be use for making disparity maps on the basis of a different matching algorithm. The aim of this research will be discovering the most beneficial combinations of matching algorithms.

## References


Agrawal, M., Davis, L., 2002. Trinocular stereo using shortest paths and the ordering constraint. Int. J. Comput. Vision 47, 43–50. http://dx.doi.org/10.1023/A:1017478504047.

Bac, C., Hemming, J., van Henten, E., 2013. Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper. Comput. Electron. Agric. 96, 148–162. http://dx.doi.org/10.1016/j.compag.2013.05.004.

Baeten, J., Donné, K., Boedrij, S., Beckers, W., Claesen, E., 2008. Autonomous fruit picking machine: a robotic apple harvester. In: Laugier, C., Siegwart, R. (Eds.), Field and Service Robotics, Springer Tracts in Advanced Robotics, vol. 42. Springer, Berlin/Heidelberg, pp. 531–539. http://dx.doi.org/10.1007/978-3-540-75404-6_51.

Belforte, G., Deboli, R., Gay, P., Piccarolo, P., Aimonino, D.R., 2006. Robot design and testing for greenhouse applications. Biosyst. Eng. 95, 309–321. http://dx.doi.org/10.1016/j.biosystemseng.2006.07.004.

Besag, J., 1986. On the statistical analysis of dirty pictures. J. Roy. Stat. Soc. Ser. B (Methodol.) 48, 259–302. http://dx.doi.org/10.2307/2345426.

Birchfield, S., Tomasi, C., 1998. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Trans. Pattern Anal. Mach. Intell. 20, 401–406. http://dx.doi.org/10.1109/34.677269.

Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. 26, 1124–1137. http://dx.doi.org/10.1109/TPAMI.2004.60.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23, 1222–1239. http://dx.doi.org/10.1109/34.969114.

Bradski, D.G.R., Kaehler, A., 2008. Learning Opencv. first ed. O'Reilly Media, Inc.

Chéné, Y., Rousseau, D., Lucidarme, P., Bertheloot, J., Caffier, V., Morel, P., Belin, É., Chapeau-Blondeau, F., 2012. On the use of depth camera for 3d phenotyping of entire plants. Comput. Electron. Agric. 82, 122–127. http://dx.doi.org/10.1016/j.compag.2011.12.007.

Deng, X.m., Wan, X., Zhang, Z.m., Leng, B.y., Lou, N.n., He, S., 2010. Multi-camera calibration based on opencv and multi-view registration. In: Proc. SPIE. http://dx.doi.org/10.1117/12.86335, pp. 765624–765624-6.

Domański, M., Stankiewicz, O., Wegner, K., Kurc, M., Konieczny, J., Siast, J., Stankowski, J., Ratajczak, R., Grajek, T., 2013. High efficiency 3d video coding using new tools based on view synthesis. IEEE Trans. Image Process. 22, 3517–3527.

Fusiello, A., Roberto, V., Trucco, E., 2000. Symmetric stereo with multiple windowing. Int. J. Pattern Recogn. Artif. Intell. 14, 1053–1066.

Geiger, A., Roser, M., Urtasun, R., 2011. Efficient large-scale stereo matching. In: Kimmel, R., Klette, R., Sugimoto, A. (Eds.), Computer Vision – ACCV 2010, 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8–12, 2010, Revised Selected Papers, Part I, Lecture Notes in Computer Science, vol. 6492. Springer, Berlin Heidelberg, pp. 25–38. http://dx.doi.org/10.1007/978-3-642-19315-6_3.

Grift, T., Zhang, Q., Kondo, N., Ting, K., 2008. A review of automation and robotics for the bioindustry. J. Biomechatron. Eng. 1, 37–54.

Gupta, M., Yin, Q., Nayar, S.K., 2013. Structured light in sunlight. In: 2013 IEEE International Conference on Computer Vision, pp. 545–552. http://dx.doi.org/10.1109/ICCV.2013.7.

Gurbuz, S., Kawakita, M., Ando, H., 2010. Color calibration for multi-camera imaging systems. In: 2010 4th International on Universal Communication Symposium (IUCS), pp. 201–206.

Hartley, R.I., 1999. Theory and practice of projective rectification. Int. J. Comput. Vision 35, 115–127. http://dx.doi.org/10.1023/A:1008115206617.

Hayashi, S., Shigematsu, K., Yamamoto, S., Kobayashi, K., Kohno, Y., Kamata, J., Kurita, M., 2010. Evaluation of a strawberry-harvesting robot in a field test. Biosyst. Eng. 105, 160–171. http://dx.doi.org/10.1016/j.biosystemseng.2009.09.011.

Hensler, J., Denker, K., Franz, M., Umlauf, G., 2011. Hybrid face recognition based on real-time multi-camera stereo-matching. In: Advances in Visual Computing. Lecture Notes in Computer Science, vol. 6939. Springer, Berlin Heidelberg, pp. 158–167. http://dx.doi.org/10.1007/978-3-642-24031-7_16.

van Henten, E., Hemming, J., van Tuijl, B., Kornet, J., Meuleman, J., Bontsema, J., van Os, E., 2002. An autonomous robot for harvesting cucumbers in greenhouses. Auton. Robots 13, 241–258. http://dx.doi.org/10.1023/A:1020568125418.

Hirschmuller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 30, 328–341. http://dx.doi.org/10.1109/TPAMI.2007.1166.

Hirschmuller, H., Scharstein, D., 2007. Evaluation of cost functions for stereo matching. In: IEEE Conference on Computer Vision and Pattern Recognition. CVPR '07, pp. 1–8. http://dx.doi.org/10.1109/CVPR.2007.38324.

Jang, W., Je, C., Seo, Y., Lee, S.W., 2013. Structured-light stereo: comparative analysis and integration of structured-light and active stereo for measuring dynamic shape. Opt. Lasers Eng. 51, 1255–1264.

Kaczmarek, A.L., 2015. Improving depth maps of plants by using a set of five cameras. J. Electron. Imag. 24 (2), 023018. http://dx.doi.org/10.1117/1.JEI.24.2.023018.

Kang, Y.S., Ho, Y.S., 2011. An efficient image rectification method for parallel multi-camera arrangement. IEEE Trans. Consumer Electron. 57, 1041–1048. http://dx.doi.org/10.1109/TCE.2011.6018853.

Kazmi, W., Foix, S., Alenya, G., 2012. Plant leaf imaging using time of flight camera under sunlight, shadow and room conditions. In: 2012 IEEE International

Symposium on Robotic and Sensors Environments (ROSE), pp. 192–197. http://dx.doi.org/10.1109/ROSE.2012.6402615.
Kazmi, W., Foix, S., Alenyà, G., Andersen, H.J., 2014. Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: analysis and comparison. {ISPRS} J. Photogramm. Remote Sens. 88, 128–146. http://dx.doi.org/10.1016/j.isprsjprs.2013.11.012.
Kolmogorov, V., 2006. Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1568–1583. http://dx.doi.org/10.1109/TPAMI.2006.200.
Kolmogorov, V., Zabin, R., 2004. What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26, 147–159. http://dx.doi.org/10.1109/TPAMI.2004.1262177.
Konolige, K., 1998. Small vision systems: hardware and implementation. In: Shirai, Y., Hirose, S. (Eds.), Robotics Research. Springer, London, pp. 203–212. http://dx.doi.org/10.1007/978-1-4471-1580-9_1.
Kosov, S., Thormählen, T., Seidel, H.P., 2009. Accurate real-time disparity estimation with variational methods. In: Advances in Visual Computing. Lecture Notes in Computer Science, vol. 5875. Springer, Berlin Heidelberg, pp. 796–807. http://dx.doi.org/10.1007/978-3-642-10331-5_74.
Kurillo, G., Baker, H., Li, Z., Bajcsy, R., 2013. Geometric and color calibration of multiview panoramic cameras for life-size 3d immersive video. In: 2013 International Conference on 3D Vision – 3DV 2013, pp. 374–381. http://dx.doi.org/10.1109/3DV.2013.5.
Li, K., Dai, Q., Xu, W., 2011. Collaborative color calibration for multi-camera systems. Signal Process.: Image Commun. 26, 48–60. http://dx.doi.org/10.1016/j.image.2010.11.004.
Li, P., heon Lee, S., Hsu, H.Y., 2011. Review on fruit harvesting method for potential use of automatic fruit harvesting systems. Procedia Eng. 23, 351–366. http://dx.doi.org/10.1016/j.proeng.2011.11.2514. {PEEA} 2011.
Matusik, W., Pfister, H., 2004. 3d tv: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. In: ACM SIGGRAPH 2004 Papers. ACM, New York, NY, USA, pp. 814–824. http://dx.doi.org/10.1145/1186562.101580.
Mehta, S., Burks, T., 2014. Vision-based control of robotic manipulator for citrus harvesting. Comput. Electron. Agric. 102, 146–158. http://dx.doi.org/10.1016/j.compag.2014.01.003.
Murakami, N., Ito, A., Will, J.D., Steffen, M., Inoue, K., Kita, K., Miyaura, S., 2008. Development of a teleoperation system for agricultural vehicles. Comput. Electron. Agric. 63, 81–88. http://dx.doi.org/10.1016/j.compag.2008.01.015 (special issue on bio-robotics).
Muscato, G., Prestifilippo, M., Abbate, N., Rizzuto, I., 2005. A prototype of an orange picking robot: past history, the new robot and experimental results. Ind. Robot: Int. J. 32, 128–138. http://dx.doi.org/10.1108/01439910510582255.
Nanda, H., Cutler, R., 2001. Practical calibrations for a real-time digital omnidirectional camera. In: Proceedings of CVPR, Technical Sketch, pp. 1–4.
Nielsen, M., Andersen, H., Slaughter, D., Granum, E., 2007. Ground truth evaluation of computer vision based 3d reconstruction of synthesized and real plant images. Precis. Agric. 8, 49–62. http://dx.doi.org/10.1007/s11119-006-9028-3.
Okutomi, M., Kanade, T., 1993. A multiple-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. 15, 353–363. http://dx.doi.org/10.1109/34.206955.
Park, J.I., Inoue, S., 1998. Acquisition of sharp depth map from multiple cameras. Signal Process.: Image Commun. 14, 7–19. http://dx.doi.org/10.1016/S0923-5965(98)00025-3.
Plebe, A., Grasso, G., 2001. Localization of spherical fruits for robotic harvesting. Mach. Vision Appl. 13, 70–79. http://dx.doi.org/10.1007/PL00013271.
Qingchun, F., Xiu, W., Wengang, Z., Quan, Q., Kai, J., 2012. New strawberry harvesting robot for elevated-trough culture. Int. J. Agric. Biol. Eng., 1–8
Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J., Kang, S.B., 2006. Image-based plant modeling. ACM Trans. Graph. 25, 599–604. http://dx.doi.org/10.1145/1141911.1141929.
Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P., 2014. High-resolution Stereo Datasets with Subpixel-accurate Ground Truth. Springer International Publishing, Cham, pp. 31–42. http://dx.doi.org/10.1007/978-3-319-11752-2_3.
Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision 47, 7–42. http://dx.doi.org/10.1023/A:1014573219977 (microsoft Research Technical Report MSR-TR-2001-81, November 2001).
Sun, C., 2003. Uncalibrated three-view image rectification. Image Vision Comput. 21, 259–269. http://dx.doi.org/10.1016/S0262-8856(02)00157-9.
Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C., 2008. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1068–1080. http://dx.doi.org/10.1109/TPAMI.2007.70844.
Tan, P., Zeng, G., Wang, J., Kang, S.B., Quan, L., 2007. Image-based tree modeling. ACM Trans. Graph. 26. http://dx.doi.org/10.1145/1276377.1276486.
Tanigaki, K., Fujiura, T., Akase, A., Imagawa, J., 2008. Cherry-harvesting robot. Comput. Electron. Agric. 63, 65–72. http://dx.doi.org/10.1016/j.compag.2008.01.018 (special issue on bio-robotics).
Tappen, M., Freeman, W., 2003. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. Proceedings. Ninth IEEE International Conference on Computer Vision, vol. 2, pp. 900–906. http://dx.doi.org/10.1109/ICCV.2003.1238444.
Wainwright, M., Jaakkola, T., Willsky, A., 2005. Map estimation via agreement on trees: message-passing and linear programming. IEEE Trans. Inform. Theory 51, 3697–3717. http://dx.doi.org/10.1109/TIT.2005.856938.
Wilburn, B., Joshi, N., Vaish, V., Talvala, E.V., Antunez, E., Barth, A., Adams, A., Horowitz, M., Levoy, M., 2005. High performance imaging using large camera arrays. In: ACM SIGGRAPH 2005 Papers. ACM, New York, NY, USA, pp. 765–776. http://dx.doi.org/10.1145/1186822.107325.
Williamson, T., Thorpe, C., 1999. A trinocular stereo system for highway obstacle detection. Proceedings. 1999 IEEE International Conference on Robotics and Automation, vol. 3, pp. 2267–2273. http://dx.doi.org/10.1109/ROBOT.1999.770443.
Xiang, R., Ying, Y., Jiang, H., Peng, Y., 2010. Three-dimensional location of tomato based on binocular stereo vision for tomato harvesting robot. In: Proc. SPIE. http://dx.doi.org/10.1117/12.86693, pp. 76582Z–76582Z-7.
Yang, J., Ding, Z., Guo, F., Wang, H., 2014a. Multiview image rectification algorithm for parallel camera arrays. J. Electron. Imag. 23, 033001. http://dx.doi.org/10.1117/1.JEI.23.3.033001.
Yang, J., Guo, F., Wang, H., Ding, Z., 2014b. A multi-view image rectification algorithm for matrix camera arrangement. Artif. Intell. Res., Sciedu Press 3, 18–29. http://dx.doi.org/10.5430/air.v3n1p18.
Yang, L., Dickinson, J., Wu, Q., Lang, S., 2007. A fruit recognition method for automatic harvesting. In: 14th International Conference on Mechatronics and Machine Vision in Practice, 2007. M2VIP 2007, pp. 152–157. http://dx.doi.org/10.1109/MMVIP.2007.443073.
Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22, 1330–1334. http://dx.doi.org/10.1109/34.888718.