Webcam for Stereoscopic Video Conversation

Radoslaw Hofman, Tomasz Zielonka

***Abstract*— In this paper the authors present an innovative approach which may be used for stereoscopic video conversation using Internet communicators. The main idea is software image processing from two differing single lens cameras, using advanced imaging techniques to obtain stereoscopic video stream, which may be presented on a 3D display. The major task is to eliminate the differences between streams, associated with the use of two, non-homogenic cameras. This solution brings fully automatic product which is ready for recalibrations.**

***Index Terms*—Internet Stereoscopic Video, Advanced Image Processing**

I. Introduction

3D video content presentation in recent years had become one of the most popular trend in video projection technology. Its growing support in cinemas, TV, computer screens etc. not excluding VR goggles allow anyone (even without dedicated device) to experience the deepness of the scene. However, despite the high popularity of 3D technology it is still not used widely for video capturing or communication using Internet communicators. The reason is quite obvious: there is a lot of effort to produce a 3D camera, therefore such devices are not easily accessible.

The problem with 3D cameras is in their construction - all the imperfections of the image due to inaccurate construction of the image recording device, differences in lens, electronic parts etc. make it difficult to manufacture a good quality device. When cameras are not identical (they record e.g. color of the image in different way) and are not perfectly positioned relative to each other – captured 3D image will not be projected correctly. The construction of the stereoscopic camera is complicated due to differences in the components used in each optical set. Typically, a single camera consists of a lens and a silicon transducer which encodes the picture based on rays of light coming to each of the transducer's cell. If two optical systems will be compared, there are noticeable differences: lenses are not identical – each one adds other distortion to the image and they are not aligned parallel to their main axes. Additionally two silicon transducers register same colors as different sets of RGB values. These problems establish a set of complex technical requirements to be fulfilled by 3D camera device, which makes their production economically inefficient for a wide consumer market.

Additional aspect is traditional image processing software which is resource intensive and requires large computational power.

The authors of this paper wanted to meet these challenges and developed the solution that minimizes the impact of construction imperfections of cameras and takes less resource computing performance compared to traditional programs implementing 3D image processing. This software can cooperate with any non-identical devices, even using built-in laptop camera with additional webcam as stereoscopic pair, what greatly expands the functionality of the software.

Manuscript received on 2015.11.29, revised on 2016.01.01.

The paper presents the research results from the innovative project Innovative stereoscopic webcam with dedicated software for image processing prototype construction - the project co-founded by European Union (project number: UDA-POIG.01.04.00-30-018/10-03, realization period: 02.2012-06.2014). European funds for the development of innovative economy (Fundusze Europejskie - dla rozwoju innowacyjnej gospodarki).

Radoslaw Hofman 3D Vision sp. z o.o., ul. Polna 3/5, 60-535 Poznan, Poland (radoslaw.hofman@3d-vision.pl)
Tomasz Zielonka 3D Vision sp. z o.o., ul. Polna 3/5, 60-535 Poznan, Poland (tomasz.zielonka@3d-vision.pl)

II. Devices on the Market

A. High quality cameras

In response to the rapidly growing demands in the 3D industry, camera manufacturers had to build a highly specialized camera for recording high quality 3D content. For this purpose they have designed cameras with advanced optics, fast graphic processors and great attention to manufacture detail, required by proper stereoscopic image acquisition. An example of such device may be the Sony PMW-TD300 dedicated to the high-quality production presented in Figure 1. This camera is characterized by interchangeable lens and the ability to change focal length for zooming. The manufacture quality of such a device is at very high level resulting in a product, that does not require calibration [1]. Recorded material is high resolution and processing of such image is performed by a specialized software inside camera. An important feature of such complete solutions is the fixed and unchangeable arrangement of two lenses inside the device. The user is unable to move or change the spacing or position of optical sets, and therefore the relative position of these sets is known. It is also known, that vertical displacement of two images is small, and the horizontal offset is approximately constant. In addition, images obtained with these cameras have very similar color characteristics, often not requiring correction - recording lenses of stereoscopic video simply write two images on the same matrix. Such systems generally can be calibrated only once. Such cameras are large, bulky, and very expensive. However, video footage obtained from this type of camera could be easily processed by the software described in this paper.

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Figure 1. High end Sony PMW-TD300 stereoscopic camera (source: sony.com)

B. Web cams as additional devices

Another type of stereoscopic cameras available on the market are standalone devices mounted on the computer screen. They are characterized by lower manufacturing accuracy which justifies lower price. An example of such a camera can be Minoru 3D shown in Figure 2. This stereoscopic camera operates as two low-quality cameras combined in one device forming a stereoscopic pair. The location of the lenses is not as accurate as in the previously described solution. The manufacturer include software with the camera which creates an image that is anaglyph so blue-red glasses are required to be able to see the 3D effect. These devices are equipped with sensors recording a low resolution image resulting in low quality of transmitted video. This is due to the fact that low-resolution image is easier to calibrate by included software. This solution is very simple and does not meet the expectations set by users and producers of video instant messaging applications as the color of the image is changed due to anaglyph type of presentation. In addition, image resolution greatly differs from the commercially available standard 2D webcams.

The problem of differences between two optical sets used to build such cameras and unknown information about their actual relative position is usually neglected. It is assumed that the observer (human) will adjust their way of observation to the differing images in anaglyph, so the main task for the producer is to manufacture the device, where differences between optical sets are not too significant. However such 3D material may not be used for any kind of automated depth analysis, as the differences between two views are generally unknown

Figure 2. Minoru 3D webcam (source www.minoru3d.com)

Humans' abilities to adjust the way of perceiving anaglyphs (compensate differences) and low cost were ones of the reasons for such a technology to be developed. In the past, the manufacturers of built-in monitor cameras have also made attempts to build 3D devices. An example might be a camera made by Sharp illustrated in Figure 3. These devices, however, did not achieve the expected success, among other things due to some limitations: the producer aimed to know the relative position of the optical sets, therefore the device had to be small – in this case the lens spacing was so low, that the stereoscopic effect was hardly noticeable. On the other hand the cost of such device was significant, as the producer was aiming to have identical lens, identical transducers (made from the same silicon wafer) etc.

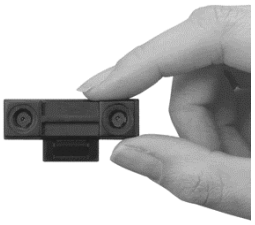


Figure 3 Module of 3D stereoscopic camera (source Sharp-world)

C. The missing part

Besides the two examples of obtaining stereoscopic footage outlined above, the authors were analyzing an additional solution which was not yet on the market: the possibility to use two different cameras as stereoscopic pair. For example one could use a built-in monitor camera along with second 2D camera creating a stereoscopic pair. Or the laptop / monitor producer could simply put two standard cameras into the monitor housing (with spacing above 7,5 cm). This solution could be simple, quick, cheap and could be implemented on any computer without problem. The idea is presented in Figure 4.

The cameras which are built into laptops' cover or monitors currently available on the market are rather good quality. Therefore using such devices for stereoscopic pairs would also be attractive in terms of resolution of the captured stereoscopic material.

Figure 4. Additional webcam attached to the built in monitor webcam

However, such combination of two cameras raises a number of issues and problems that must be solved so the combination into stereoscopic pair could work properly. The most important of these is the approach of compensation of differences in position, size, colors, geometry and distortions etc.

Cameras wouldn't be (in general) positioned parallel to each other or even in one horizontal line. It should also be assumed that they will record the scene from different angles. The distance between two camera lenses and recorded object is also variable. The image sensors and lenses are from different manufacturers and they record color in two various ways with various optical distortions to

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the geometry on the image. They may also differ in terms of resolution. A number of these problems means that such a solution is not that easy to implement.

The authors of this paper wanted to meet the requirements set out above. The idea was to create software which algorithm would eliminate the need to perform tedious and accurate calibration of cameras shifting the responsibility to compensate differences to software (including also the differences in color and image resolution and proper reaction to spontaneous re-configuration of the stereoscopic pair – if during the usage time the user will move one of the cameras to the new position or would replace one of the cameras with a new one).

III. Proposed solution

The solution proposed in this paper offers 3D scene reconstruction using two non-homogenous single-lens devices. The acquired images are used by the software to create a single video stream with spatial scene. This process is shown step by step in Figure 5.

Fig. 5. General steps of the proposed approach

The first step is understood as the acquisition of video footage from two cameras at once. The proposed solution involves the use of one camera which (for example) is built-in in computer monitor, and another camera (for example) attached to the monitor. This setup creates stereoscopic pair which correspond respectively to the image recorded by the left and right eye of a human.

The next more complex step is camera calibration [1]. The calibration of cameras includes processes such as: elimination of distortion inside lens, obtaining reference points and estimation of both internal and external parameters of cameras (compare [8]).

Because the lens distortion cause problems in obtaining the correct reference points, elimination of lens distortion should be carried out in the first place while acquiring a pair of stereoscopic images. Distortions introduced by the optical lens are the most visible on the edges of the image. They cause distortion of objects - straight lines become curves. Lens distortion correction is performed for each image acquired from both cameras separately. The method is based on the elimination of distortions using the calibration table – for example a checkerboard pattern presented on Figure 6. Such a calibration object has known geometry - alternate white and black squares. It is presented in front of cameras only once after the set-up of the stereoscopic pair. With this texture, array recorded by the camera provides information about straight lines in the scene and allows to discover the effects on images produced by each camera.

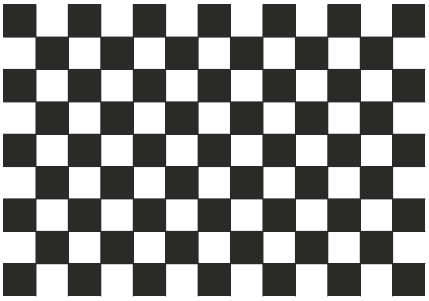


Figure 6. Calibration matrix used in the solution

Lens distortion cause that a straight line is more or less curved. This effect is more visible as it is closer to the edge of an image. The algorithm estimates the value of the parameter called distortion coefficients. Then, each pixel is transformed using those coefficients. The transformation are performed separately for radial distortion, and for tangensoidal distortion. For this method the result is a slightly smaller extracted part of the image than image registered in reality. But this is not a big inconvenience, because eliminated areas would not be suitable in a manner sufficient to be used in the system with multiple cameras.

The next step is to obtain the corresponding image points. In order to ensure proper operation of the system it is necessary to do the most accurate determination of the mutual position of the cameras. It is impossible to accurately determine their positions in advance. This would involve the need of their attachment to a rigid structure - frame. Furthermore, there would also be a need of knowledge of the internal optical characteristics of each device. To avoid these inconveniences, position of the cameras should be estimated before registering the image. For this purpose, the same calibration object is used as described above - a checkerboard pattern. Undoubtedly, it is a marker that provides the greatest number of characteristic points. This number depends of the number of black and white squares provided at its creation. In this case, characteristic points are expected, where the four neighboring squares (two blacks and two whites) touch with each other. Calibration object can be used to determine both external and internal parameters of each individual camera. However, in order to be able to determine the internal parameters of the cameras, it is necessary to meet the additional assumptions - knowing the exact dimensions of the matrix – in other words: the distance between the characteristic points on the checkerboard.

The algorithm used to compare two images and identify characteristic points used in the described system is a SIFT algorithm (Scale-Invariant Feature Transform) [5]. This algorithm is not very sensitive to change of lighting or any kinds of noise, rotation of objects, their proportions, and even a partially overshadowed exposure. The main idea of the algorithm is to extract characteristic points from recorded image. These points at further stage are used to search for similar objects in both images, making their proper alignment possible. In this method, image characteristics are not compared directly, but through descriptors - vectors representing numerically the most important information about individual areas of each image. With this approach it is possible to omit the influence of the

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

unessential characteristics of the image. For example the rotation of the cameras relative to each other, and under exposed objects due to spacing between devices. The descriptors describe the local characteristics of the image, regardless of the layout of objects in the image. Example of descriptor for a single point of the image is presented in Figure 7. The described point using a descriptor is shown as the central point in the figure.

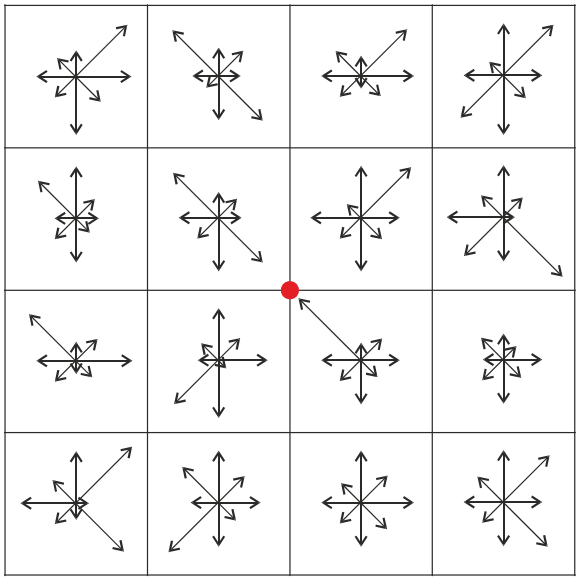


Figure 7. Visualization of descriptors for a single point.

The first step of the algorithm is to detect all the potential characteristic points in the image (compare [2, 3]). For this purpose, the input image is subject to decimation. The second stage of the algorithm is determining the location of the characteristic points. This step involves the analysis of set of potential points and rejects some of them. Points that have too low contrast are eliminated along with those that lie on the same straight line. This allows to avoid errors due to the influence of noise and any disturbance, as well as rounding errors. The next step of the SIFT algorithm is the orientation of characteristic points [5]. Intensity gradient and its orientation may be calculated for each image point. This allows to recognize the same objects (characteristic points) from different perspectives. In order to create a descriptor to the point, the histogram of gradients of neighboring points is calculated. The final step is performed separately for images from both cameras – this step allows to calculate descriptors based on previously determined characteristic points. For each point a vector which describes this point is created. After creating descriptors for all characteristic points for both images, descriptors are compared in order to find the pairs of the most similar characteristics.

SIFT algorithm is accurate enough tool for automatic determination of reference points what serves to determine the position between two cameras. An example of SIFT algorithm and extracted characteristic points is presented in Figure 8.

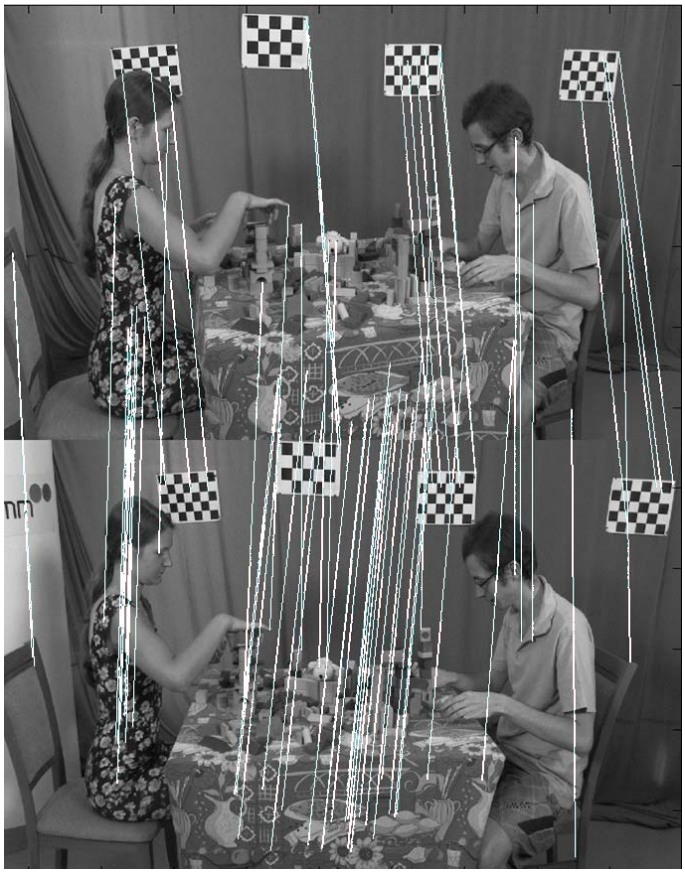


Figure 8. Operation of SIFT algorithm result

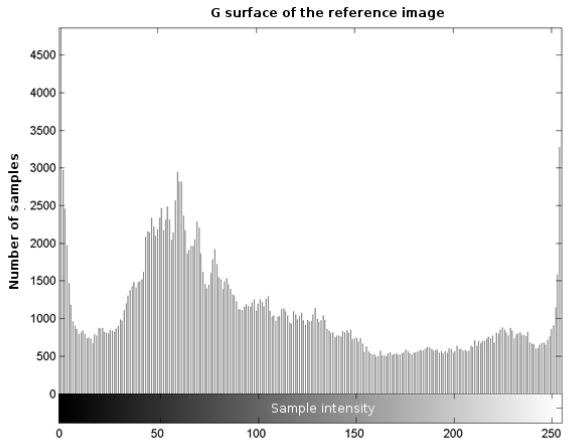
The next stage of the algorithm is focused on color unification (see [4]). Identical objects recorded by two cameras have different colors because, as mentioned above, transducers usually differ from each other even if they are manufactured on the same production line [6]. The aim of the algorithm is to eliminate these differences (including the correction of brightness, contrast, color saturation, etc.) resulting from physical differences in the structure and configuration of cameras [7]. These differences have the potential to be so deep that further image processing may be significantly hindered, or will be characterized by much less precision.

The purpose of color correction is the image processing in which these differences will be minimized. Presentation of stereoscopic images with the images that have not been calibrated in terms of color, can cause discomfort on the viewer side and lower subjective rating of the quality of the sequence. Color correction is therefore an important step for both the producer and the user. The color unification method is called histogram matching method. The aim of the histogram matching method [11] is processing a distorted image in order its histogram was the most similar to the reference image histogram. As a result of such conversion, obtained image is improved. The point values are changed in such a way, that the color spaces of two images were possibly related. Histogram matching method works on each component of RGB independently.

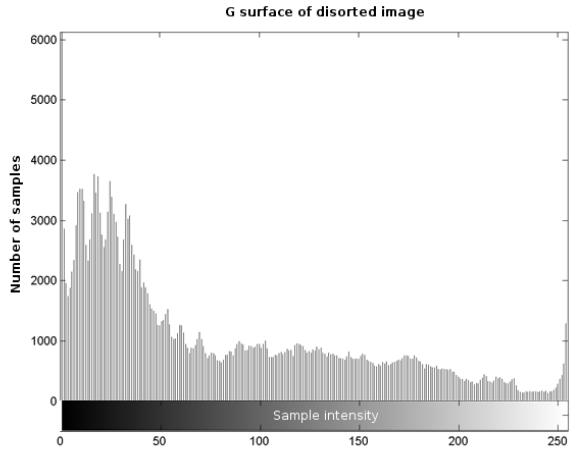
At the beginning of this step the reference image histogram must be found. Then, the cumulative histogram of the reference image has to be calculated and in the next step the distorted image histogram along with cumulative histogram of distorted image has to be prepared. As the

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

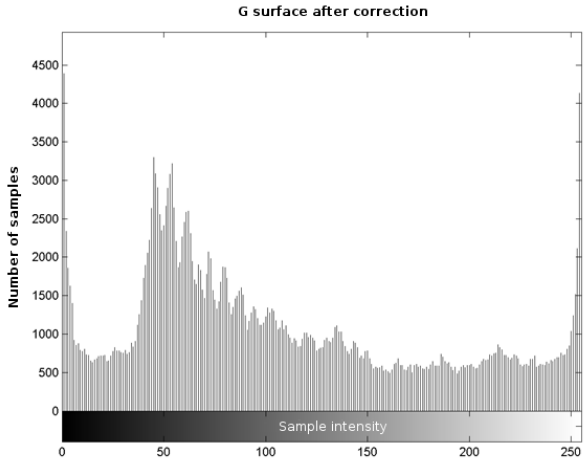
result a function, based on the histogram and the cumulative histogram of distorted image and reference image is prepared to be used on video streams. This function assigns values of samples of distorted image sample values of the reference image. The last step is to convert the whole distorted image in accordance with the function. As a result of this transformation, there is obtained the correct image. The result of this process is presented on Figure 9.



Histogram component Green of reference image.



Histogram component Green of distorted image.



Histogram component Green of the image after correction.

Figure 9. Process of histogram matching method

In general it also have to be assumed that cameras of different construction and coming from different manufacturers do not have the same resolution (compare [10]). In the case of histogram matching method this assumption does not change the approach, because each value of the obtained histogram is divided by the number of samples in the image. Therefore histograms represent relative content of samples of the given color. The proposed method does not require undertaking any additional calculations and is suitable for images of the same or different resolutions.

The last step of stereoscopic image creation is related to depth analysis [8, 9]. Using the information contained in the two video streams (it is assumed that the spatial distance between cameras is relatively constant for a sequence, so the identical points on two images are located in different places), as well as the information on the geometry distortions between the cameras, it is possible to prepare stereoscopic stream. The stereoscopic stream is formed by selecting the next lines of images from the alternately left and right cameras. The final effect is shown in Figure 10.

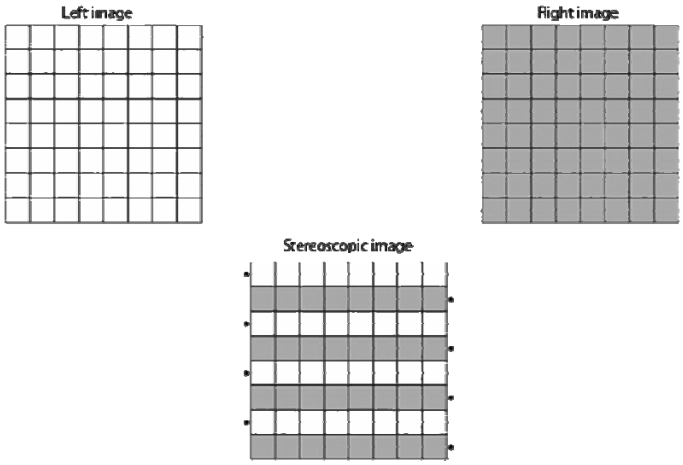


Figure 10. Process of forming stereoscopic image

IV. Experimental results

The approach presented in this paper was designed in such a way, that it is able to utilize simple devices available on the market without the need of the construction of a specialized device which is the 3D camera. This approach has been tested in many cases using a variety of popular low-cost 2D cameras connected as stereoscopic pairs.

The main goal of the authors was to create algorithm that would be able to be used by manufacturers of personal computers, tablets, smartphones, and VR goggles without having knowledge of 3D imaging techniques.

The level of quality of the reconstructed 3D scenes depends mainly of the amount of light filling the scene and the resolution of the cameras used to generate an image, however these factors are not unusual in terms of influence of the recorded stream quality. The proposed algorithm gives promising results in terms of ease of usage and the quality of reproduced spatial scenes. The authors believe that quality is fully satisfactory for casual conversations using video chatting programs on a variety of hardware

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

platforms. Algorithm does not require any hardware calibration and allows the user to combine into stereoscopic pair even non-homogenic cameras without precise positioning.

V. Conclusions

The algorithm presented in this document has been tested for different combinations of 2D cameras connected in pairs - creating a 3D device to record stereoscopic images. It is easy to use and, as previously mentioned, does not require any hardware calibration of the devices. Assumption of the authors was to create software that does not require specialist knowledge about 3D techniques from the end-user.

References

[1] Cyganek B (2013) Object Detection and Recognition in Digital Images: Theory and Practice, Wiley
[2] Fishler M., R. Bolles. „Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.” Communication of the ACM 24, 1981: 381-395.
[3] Hartley R., Zisserman A. (2003) Multiple view geometry in computer vision, Cambridge University Press
[4] Pollefeys, M., R. Koch, i L. Van Gool. „A simple and efficient rectification method for general motion.” Proc. International Conference on Computer Vision, 1999: 496-501.
[5] Lowe David G. (2004) Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 2:91-110
[6] Ilie, Adrian, i Greg Welch. „Ensuring Color Consistency across Multiple Cameras.” Proceedings of the Tenth IEEE International Conference on Computer Vision. 2005.
[7] Joshi, Neel S. Color Calibration for Arrays of Inexpensive Image Sensors. Stanford University, 2004.
[8] Brown, D. C. „Close-Range Camera Calibration.” Photogrammetric Engineering 8, vol. 37.
[9] Abdel-Aziz, Y. I., H. M. Karara. „Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry.” ASP Symposium on Close-Range Photogrammetry. Proceedings. , 1971: 1-18.
[10] Svoboda T., D. Martinec, T. Pajdla. „A Convenient Multi-Camera Self-Calibration for Virtual Environments.” Presence, vol. 14, 2005.
[11] Correal, Raúl, Gonzalo Pajares, and José Jaime Ruz. "Automatic expert system for 3D terrain reconstruction based on stereo vision and histogram matching." Expert Systems with Applications 41.4 (2014): 2043-2051.

ISBN: 978-988-14047-6-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)