Metrol. Meas. Syst., Vol. XX (2013), No. 2, pp. 229–238.

PAN POLSKA AKADEMIA NAUK

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

M&MS

AUTOMATIC CALIBRATION OF STEREOSCOPIC CAMERAS IN AN ELECTRONIC TRAVEL AID FOR THE BLIND

Paweł Pełczyński[1)], Bartosz Ostrowski[2)]

1) Łódź University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, Poland (✉ pawel.pelczynski@p.lodz.pl, +48 42 631 26 17)

2) Łódź University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, Poland (bartosz.ostrowski@p.lodz.pl)

Abstract

The article describes a technique developed for identification of extrinsic parameters of a stereovision camera system for the purpose of image rectification without the use of reference calibration objects. The goal of the presented algorithm is the determination of the mutual position of cameras, under the assumption that they can be modeled by pinhole cameras, are separated by a fixed distance and are moving through a stationary scene. The developed method was verified experimentally on image sequences of a scene with a known structure.

Keywords: stereovision, canonical setup, extrinsic parameters, image rectification, disparity.

1. Introduction

The motivation of the presented work is the development of an electronic travel aid for the blind at the Institute of Electronics of Lodz University of Technology. The system will utilize stereovision as input and spatial audio as output and allow for advanced obstacle detection and avoidance [1]. The device shown in Fig. 1. was designed to be similar to large sunglasses, to be worn on the user's head. It is equipped with two wide-angle digital cameras, sound card and inertial sensors [2]. The captured images and inertial signals are transmitted to a mobile PC via USB, where scene analysis and obstacle auditory presentation takes place.

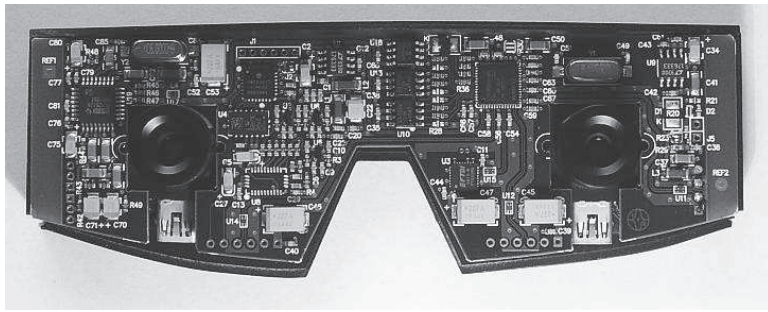

Fig. 1. A view of the hardware interface for the electronic travel aid for the blind.

To ensure correct scene reconstruction from a pair of stereovision images from a narrow-base setup, it is necessary to correct the optical distortions introduced by the lenses and compensate for the slight deviations from the intended parallel alignment of camera axes caused by imperfect housing. If the correction process succeeds, a canonical camera system

Article history: received on Oct. 23, 2012; accepted on Mar. 25, 2013; available online on Jun. 03, 2013; DOI: 10.2478/mms-2013-0020.

can be obtained [3], in which any details of the registered scene images should be at the same height y in both the left and right image ($y_L = y_R$). The difference in the horizontal coordinate x of any scene point is referred to as the disparity d [4]:

$$d = x_L - x_R . \tag{1}$$

Once the disparity is known, three-dimensional reconstruction of the observed scene is possible, by solving the following set of equations for every point:

$$X = \frac{x_L Z}{f}; Y = \frac{y_L Z}{f}; Z = \frac{fB}{d} . \tag{2}$$

where B (basis) is the distance between the optical axes of the cameras and f is the focus length of the lenses. The vector $[X,Y,Z]^T$ represents the spatial coordinates of a point in the reference frame of the left camera.

There are a number of well established methods of calibration of stereovision systems using reference objects, such as checkerboards [1]. The internal parameters of the cameras (such as the geometric distortion coefficients) and extrinsic parameters (defining the mutual arrangement of cameras) allow effective correction and rectification of images by mapping the resulting image to the output of the camera. To ensure stability of the reconstruction conditions, rigid mounting of the cameras is required, as even a minimal change of their relative positions can cause errors in determination of the disparity and, consequently, the Z component that is inversely proportional to it. The problem becomes particularly important in the case of mobile stereovision systems and augmented reality applications. The stereovision system shown in Fig. 1. has been mounted on a laminate, which is also the electronic control system for image acquisition [2]. The system has been calibrated using a standard procedure using a reference checkerboard. Despite the fairly rigid mounting of the cameras (Fig. 1.) during the test and during modifications in the system, there were occurrences of slight changes in the mutual orientation of the cameras and drops in the accuracy of disparity calculations. The internal parameters of the cameras did not change nor did the distance between them, as confirmed by the observation of parameters obtained during repeated calibrations.

The special nature of the developed application, namely supporting the mobility of blind persons, places stringent requirements on the design of the stereovision system. It must be lightweight and highly miniaturized, which does not allow very high rigidity of the camera mounting. Additionally, there is always the danger of mishandling or impact that can permanently change the mutual orientation of the cameras. Calibration procedures that could be repeatedly performed by a blind user are not possible, so there is a need to develop methods allowing to automatically adjust the parameters of the stereovision system.

2. Algorithm for automatic calibration of extrinsic camera parameters

In recent years, a number of techniques have been developed that allow calibration of stereovision systems that do not require registration of reference objects with known spatial structure. Authors of [5] propose a continuous calibration algorithm which combines both the epipolar constraint for instantaneous measurements and recursive bundle adjustment (it accumulates information from spatial temporal correspondence trajectories over time). Unfortunately, this technique suffers from high computational complexity. Other approaches, such as the one described in [6], require the identification of distant scene elements with simple structure (such as lines).

In the presented application in addition to the calibration, the system continuously calculates a dense disparity map, which places a significant time constraint on the calibration algorithm. To enable automatic calibration of the stereovision setup without a reference object, it is assumed that the cameras remain at a fixed distance from each other and the parameters of the geometric distortions are immutable. Also for calibration the user must move through a scene in which it is possible to register remote objects for which the disparity tends to zero. The left camera in the stereovision setup (Fig. 2.) is used as the reference camera. The spatial points in the system of the left camera are marked as $\boldsymbol{P}=[X,Y,Z]^T$, and the spatial points in the right camera system are referred to as $\boldsymbol{P}_R=[X_R,Y_R,Z_R]^T$. The relationship between the coordinates in the camera reference system and the coordinates of the other camera system can be determined by the formula:

$$\boldsymbol{P}=\boldsymbol{R}\cdot\boldsymbol{P}_R+\boldsymbol{T}, \tag{3}$$

where $\boldsymbol{R}$=[*r1 r2 r3*] is the rotation matrix and $\boldsymbol{T}$ is the translation vector between the left and right cameras. Vectors *r1, r2, r3* in $\boldsymbol{R}$ matrix are three element orthonormal vectors.

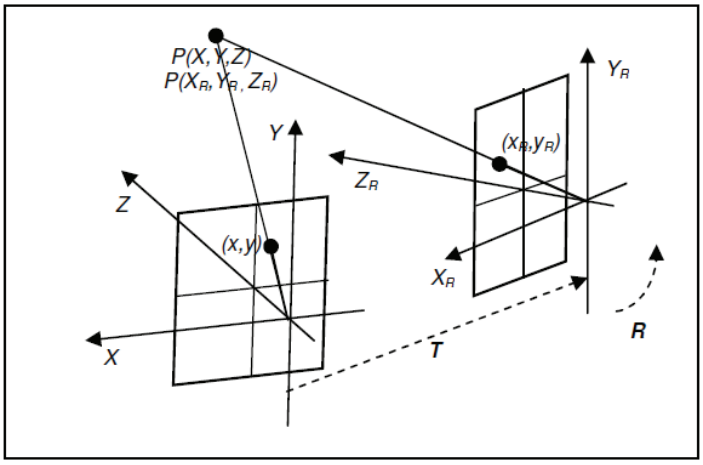

Fig. 2. Image sensors of the stereoscopic setup and the reference frames associated with them.

It is assumed that the camera mount is stable enough to prevent relative translational movement, so the translation vector is constant and equal to $\boldsymbol{T}$=[-*B*, 0, 0] ^{T}. For small rotation angles the rotation matrix can be approximated by a simplified form [7], which allows to write equation (3) as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & \gamma & -\beta \\ -\gamma & 1 & \alpha \\ \beta & -\alpha & 1 \end{bmatrix} \begin{bmatrix} X_R \\ Y_R \\ Z_R \end{bmatrix} + \begin{bmatrix} -B \\ 0 \\ 0 \end{bmatrix}. \tag{4}$$

Angles: α, β and γ are the angles of camera rotation around the *X*, *Y* and *Z* axes. For sake of simplicity, the camera focal length is given the normalized value of $f = 1$. The image coordinates of the left and right image of the point $\boldsymbol{P}$ are given by:

$$x=\frac{X}{Z};y=\frac{Y}{Z};x_R=\frac{X_R}{Z_R};y_R=\frac{Y_R}{Z_R}. \tag{5}$$

Transforming (3) it is possible to write the coordinates of point ***P*** in the right camera image as a function of its coordinates in the left camera image:

$$\begin{cases} x_R = \dfrac{X_R}{Z_R} = \dfrac{X + B - \gamma Y + \beta Z}{-\beta X - \beta B + \alpha Y + Z} = \dfrac{xZ + B - \gamma yZ + \beta Z}{-\beta xZ - \beta B + \alpha yZ + Z} \\ y_R = \dfrac{Y_R}{Z_R} = \dfrac{\gamma X + \gamma B + Y - \alpha Z}{-\beta X - \beta B + \alpha Y + Z} = \dfrac{\gamma xZ + \gamma B + yZ - \alpha Z}{-\beta xZ - \beta B + \alpha yZ + Z} \end{cases}. \quad (6)$$

Distant points satisfy the condition $Z>>B$, which allows to disregard the B components in (6), making the image coordinates independent of the actual position in space:

$$\begin{cases} x_R = \dfrac{x + \dfrac{B}{Z} - \gamma y + \beta}{-\beta x - \dfrac{\beta B}{Z} + \alpha y + 1} \approx \dfrac{x - \gamma y + \beta}{-\beta x + \alpha y + 1} \\ y_R = \dfrac{\gamma x + \dfrac{\gamma B}{Z} + y - \alpha}{-\beta x - \dfrac{\beta B}{Z} + \alpha y + 1} \approx \dfrac{\gamma x + y - \alpha}{-\beta x + \alpha y + 1} \end{cases} ; Z >> B. \quad (7)$$

The above equations can be transformed and rewritten in matrix form:

$$\begin{bmatrix} x - x_R \\ y - y_R \end{bmatrix} = \begin{bmatrix} yx_R & -(xx_R + 1) & y \\ (yy_R + 1) & -xy_R & -x \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}. \quad (8)$$

The equation (8) should be solved for a large number of distant scene points. The linear system with the searched angles as unknowns can be solved quickly and non-iteratively, similarly to equations for motion estimation in [7]. Once the rotation matrix is known, the images can be rectified using a standard algorithm described in [8].

3. Searching for distant scene points

The algorithm for detecting distant scene points consists of the following operations:

- Searching for characteristic points in the left image using a Harris corner detector,
- Determination of the corner movement between consecutive images, verified with angular measurements from an electronic MG1101B gyroscope [9],
- Determination of disparity for the corner points,
- Iterative model refinement using (8) and selection of corners basing on correspondence to the model using the RANSAC algorithm.

Verification of the image data with gyroscope measurements allowed the rejection of points located too close to the system. Movement of distant static scene points arises only from the rotation of the stereovision system and can be well-predicted by measuring the angles of rotation with an independent sensor – the gyroscope. For $f = 1$ and small rotation angles of the camera between successive images the dependency can be described as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} xy & -\left(1+x^2\right) & y \\ \left(1+y^2\right) & -xy & -x \end{bmatrix} \begin{bmatrix} \omega_X \cdot \Delta t \\ \omega_Y \cdot \Delta t \\ \omega_Z \cdot \Delta t \end{bmatrix}, \tag{9}$$

where:
u ,v – image movement vectors in the x and y directions,
ω_X, ω_Y, ω_Z – angular velocity components of the stereovision system,
Δt – time between consecutive images captured by the stereo camera.

Points that do not meet the calculated motion model are rejected. This approach gives good results while continuously moving through a scene, which is a typical situation for a support system for a blind person. Mapping (7) of pixels in the left image to the right one is independent of the movement, dependent only on the mutual orientation of the cameras. In this situation, it was possible to accumulate data of the points positions for a few seconds before running the algorithm for rotation angle estimation. The accumulation procedure consists of the following operations, performed for each image in the sequence:

- detection of new characteristic points - corners,
- adding the detected corners to the set of tracked points,
- verification of the points' movement with the model (9): increasing the certainty that distant points were identified correctly and rejecting points which do not meet the model,
- moving the points with high enough certainty to the set of accumulated points.

The measure of certainty of a scene point being distant is expressed as the number of frames its motion is compatible with the model (9). Seemingly, this time should be as long as possible, but an excessive increase of the threshold results in too many points having the opportunity to leave the field of view of the cameras or get occluded by scene objects. Optimal results for test scenes were obtained with a threshold between 0.5s and 1s. Compliance with the motion model is estimated basing on the distance between the expected and the actual displacement vector for a corner point in two consecutive image frames. Due to the accuracy limits of the angular velocity measurement by the gyroscope, the threshold distance cannot be too small. With the MG1101B gyroscope used in the system, it was necessary to set the threshold to a distance of 3 pixels. After 1000 points were accumulated, the angle estimation algorithm was launched iteratively fitting the model to follow and reject points that generated large errors. For this purpose a modified RANSAC algorithm was used, which specified the number of iterations for random input data selection and model fitting iterations to find the correct data. This approach allows to measure the time required to perform the calculations with good accuracy, as it becomes independent of the input data. A block diagram of the algorithm is shown in Fig. 3.

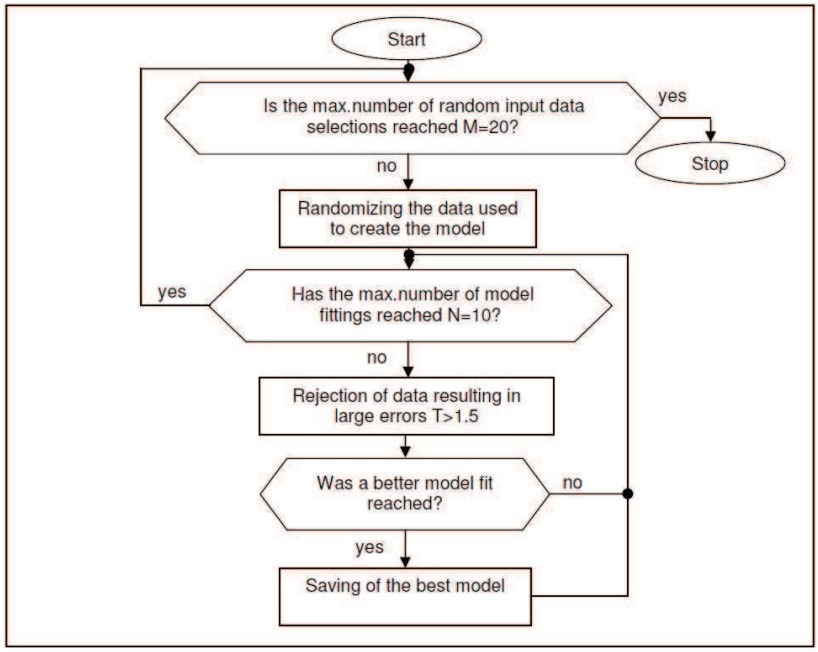

Fig. 3. Block diagram of the modified RANSAC algorithm used to select corner points that are classified as distant from the cameras.

Basing on a series of sample trials the following algorithm parameters were set:
- number of iterations of random input data selection M = 20,
- number of iterations for model fitting N = 10,
- error threshold, above which a data point is rejected T = 1.5pix.

4. Experimental verification of the camera calibration technique

In order to verify the correctness of the proposed approach a synthetic data set was prepared, containing image sequences of a virtual scene and modeled signals from a triaxial gyroscope. Image synthesis was carried out in Pov-Ray after designing a virtual trajectory for a stereoscopic camera. To simulate real motion, the mobile camera system underwent curvilinear movements with variable speeds and variable rotation. The modeled stereo system was the canonical setup with 0.1m base and 60 deg view angle cameras. The rotation matrix for such a system is the unit matrix.

Fig. 4a) shows characteristic image points for which the movement was verified by the gyroscope (9). It can be seen that points belonging to nearby objects were discarded. The number of points tracked and accumulated as a function of the frame number is shown in Fig. 4b). The first value depends on the number of corners detected in successive images, as well as the number of points removed from the set either due to non-compliance with the motion model or reaching the maximum number of tracked points. The number of points accumulated in 0.7s is zero, as initially no points with sufficient certainty of belonging to distant objects were available. Then the rate of accumulation of successive corners depends both on the movement of the cameras and the shape of the observed scene.

a) b)

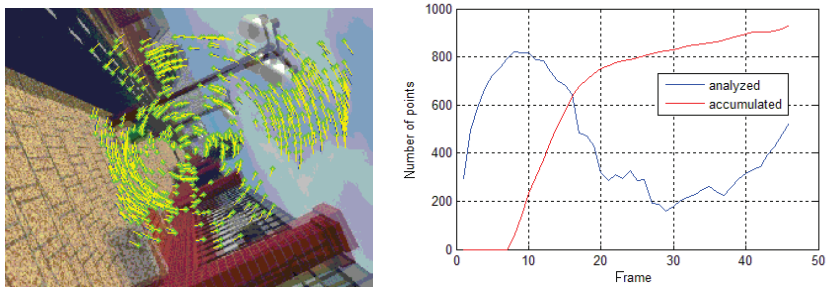

Fig. 4. a) Tracked characteristic points, b) Sample plot of the number of analyzed and accumulated points as a function of frame number.

After accumulating a thousand points, the procedure for the determination of the model (8) was launched. Although the solution of an overdetermined equation system boils down to determining the matrix (generalized inverse, pseudoinverse) to the matrix of known factors, it is prone to inaccuracy due to points belonging to nearby objects that were not removed at the stage of pre-selection. The RANSAC algorithm effectively eliminates these points, well improving the estimation of the rotation matrix of the right camera. Fig. 5.a) shows a set of accumulated points with a high probability of belonging to distant objects and their disparity vectors. Some of them were wrongly accumulated, despite their disparity being too large. The use of an iterative algorithm to determine the model, allowed to correctly remove points introducing large errors, which can be seen in Fig. 5.b).

a) b)

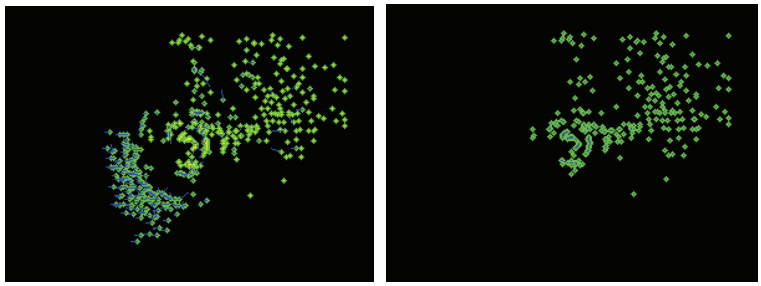

Fig. 5. a) Accumulated characteristic points with their disparity vectors, b) Points remaining after selection using the RANSAC algorithm.

Final values for the rotation vectors in the test sequence were equal to:

$$\alpha = -4.7380751\text{e}-005; \quad \beta = -1.6540943\text{e}-004; \quad \gamma = 3.2302207\text{e}-004. \tag{10}$$

As for the simulated canonical setup the values should be zero, the numbers above can be treated as the error. The matrix determined using (4) has been subjected to orthonormalization, which finally gave the following form:

$$\boldsymbol{R}_{\text{determined}} = \begin{bmatrix} 9.9999993\text{e}-001 & 3.2301987\text{e}-004 & 1.6539414\text{e}-004 \\ -3.2301205\text{e}-004 & 9.9999995\text{e}-001 & -4.7334175\text{e}-005 \\ -1.6540942\text{e}-004 & 4.7280747\text{e}-005 & 9.9999999\text{e}-001 \end{bmatrix}. \tag{11}$$

The purpose of estimating the right camera rotation parameters is to rectify the image and provide a canonical stereoscopic view in both cameras. This means that very distant objects should appear in identical positions in the left and right images. The displacement of the image points determined from (7) on the basis of the values of the angles in (10) was up to 0.1653 pixels on the *x*-axis and 0.0766 pixels on the *y*-axis. These values indicate a very low level of the estimation error for the camera rotation angles, allowing accurate determination of disparity from rectified images [10].

The next test was a quantitative assessment of the effectiveness of the proposed method on a real scene. The algorithm analyzed recorded image sequences of a scene using the stereo setup (Fig. 1.) with the following parameters: B = 0.08m, focal length 843 pixels, image size 640x480 pixels. The sequence consisted of 200 images recorded with a 15fps frame rate while walking through a scene with typical walking speed. The extrinsic parameters of the stereoscopic cameras were first intentionally modified, leading to errors in disparity calculations dependent on the shape of scene elements and their location in the image (Fig. 6.). The accumulation of points from distant objects was difficult because of the actual design of the scene, but the farthest points were correctly identified, as can be seen in Fig. 7.a). The results of the estimation of the rotation matrix elements were output to a file. After image rectification disparity calculations were correct. The vectors for the selected characteristic points in Fig. 7.b) have correct (purely horizontal) directions.

Quantitative measurements of the correctness of calibration and rectification of the stereo cameras were performed. For this purpose, measurements of the extrinsic parameters were performed with the use of a calibration checkerboard and were compared with rotation matrices determined by the developed algorithm.

a) b)

Fig. 6. Examples of 3D scene images with feature points and their disparity vectors without proper image rectification.

a) b)

Fig. 7. a) Feature points selected for rotation angle estimation,
b) Disparity vectors of feature points after image rectification.

The rotation matrix determined through standard calibration using a reference board was as follows:

$$\boldsymbol{R}_{reference} = \begin{bmatrix} 0.9984626 & -0.0278903 & -0.0479015 \\ 0.0267948 & 0.9993679 & -0.0233623 \\ 0.0485228 & 0.0220429 & 0.9985788 \end{bmatrix}. \tag{12}$$

The angles of rotation determined using the proposed method for dynamic calibration were equal to:

$$\alpha = -2.3379073\text{e}-002; \quad \beta = 4.7892524\text{e}-002; \quad \gamma = -2.7890164\text{e}-002. \tag{13}$$

The rotation matrix for the right camera is thus:

$$\boldsymbol{R}_{estimated} = \begin{bmatrix} 9.9846775\text{e}-001 & -2.8987179\text{e}-002 & -4.7136898\text{e}-002 \\ 2.7847366\text{e}-002 & 9.9930795\text{e}-001 & -2.4660548\text{e}-002 \\ 4.7819117\text{e}-002 & 2.3310123\text{e}-002 & 9.9858398\text{e}-001 \end{bmatrix}. \tag{14}$$

Errors in mapping between the left and right image points were maximally 1.0825pix and 0.9372pix for the *x* and *y* axis respectively.

The presented results demonstrate the correctness of the algorithm, proving it to be effective in the recorded scene with both close and distant objects observed by a stereovision setup worn in the form of eyeglasses. The disadvantage is that the algorithm requires continuous translational motion of the cameras for proper selection of characteristic distant scene points. There are plans to integrate the algorithm with an egomotion tracking technique [9] that could assess the suitability of consecutive frames for model building depending on the detected camera movement.

5. Conclusions

An original algorithm for estimation of extrinsic parameters of a narrow base stereovision system was developed, that did not require reference objects or human interaction. Integration of visual information with the angular velocity vector obtained from integrated gyroscopes allowed to reject data that was irrelevant to identifying the camera's misalignment angles. The correctness of the calibration results was verified quantitatively by comparison with a precise calibration procedure using a reference board. The ability to correct the extrinsic parameters of a mobile stereoscopic system during normal operation of the device for the blind will significantly increase the system's reliability and allow prolonged operation without maintenance by qualified personnel.

Acknowledgements

The presented work was financed by grant number N R02 0083 10 from the Ministry of Science and Higher Education in years 2010-2013.

References

[1] Bujacz, M., Skulimowski, P., Strumillo, P. (2012). Naviton - a prototype mobility aid for auditory presentation of 3D scenes. *Journal of Audio Engineering Society*, vol. 60, Issue 9, 696-708.

[2] B. Ostrowski, B., Pelczynski, P., Danych, R. (2011). Hardware interface vision&sound system to support the blind in the independent movement. *Przeglad Elektrotechniczny*, 10/2011, 130-132.

[3] Ma Y., Soatto S., Kosecka J., Sastry S. (2003). *An Invitation to 3D Vision: : from images to geometric transformations*. Springer-Verlag, New York, inc.

[4] Brown, M.Z., Burschka, D., Hager, G.D. (2003). Advances in computational stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.25, no. 8, 2003, 993-1008.

[5] Dang, T., Hoffmann, C., Stiller, C. (July 2009). Continuous Stereo Self-Calibration by Camera Parameter Tracking. *IEEE Transactions on Image Processing*, vol. 18, No. 7, 1536-1550.

[6] Nedevschi, S., Vancea, C., Marita, T., Graf, T. (December 2007). Online Extrinsic Parameters Calibration for Stereovision Systems Used in Far-Range Detection Vehicle Applications. *IEEE Transactions on Intelligent Transportation Systems*, ol. 8, No. 4, 651-660.

[7] Pelczynski, P., Ostrowski, B., Rzeszotarski, D. (2009). A mobile system for passive navigation in 3D scene", *VIII Electronic Conference KKE*, Darlowko Wschodnie, 07-10 June 2009, 682-687, (in Polish), Published in Elektronika (2010) R51(1), 35–37.

[8] Fusiello, A., Trucco, E., Verri, A. (2000). A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, vol. 12, 16–22.

[9] Pelczynski, P., Ostrowski, B., Rzeszotarski, D. (2012). Motion vector estimation of a stereovision camera with inertial sensors. *Metrology and Measurement Systems*, vol. XIX, No.1/2012, 141-150.

[10] Skulimowski P., Strumillo, P. (2008). Refinement of depth from stereo camera ego-motion parameters. *Electronics Letters*, vol. 44, Issue 12, 729 – 730.