An integrated thermal and visual camera system for 3D reconstruction

by D. Rzeszotarski, B. Więcek

Institute of Electronics, Technical University of Lodz, Lodz 90-924 Wolczanska 211/215, Poland*,* *dariusz.rzeszotarski@p.lodz.pl, boguslaw.wiecek@p.lodz.pl*

Abstract

In this article a new approach for 3D reconstruction of thermal scene is presented. The system consists of two cameras: thermal and visual. The reconstructed object is observed from at least two different viewpoints by the two cameras. The two images either from thermal camera or visual one constitute a stereoscopic set. By means of corresponding feature points extracted from the pair of either thermal or visual images, the relative position of the two viewpoints is evaluated. Then the images are rectified and the standard stereo reconstruction is performed on either thermal or visual images depending on the scene conditions, i.e. depending on the number and quality of feature points extracted from the thermal and visual pairs. The relative position of the two viewpoints is established only up to some scalar factor. To establish the relative position with respect to some metric value two approaches were applied: (a) based on a known distance between two points in 3D space and (b) based on an additional approximate 3D reconstruction of a given point by the thermal and visual camera pair treated as a stereoscopic set. The designed application which performs the calibration of the system and 3D scene reconstruction is described and the obtained results are presented.

1.   Introduction

Common 3D reconstruction visual systems usually consist of at least two cameras known as a stereoscopic set [1,2,3]. In order to perform a 3D reconstruction of thermal scene such visual stereoscopic systems are coupled with thermal camera [4,5,6,7,8]. It is presumed that the system is calibrated i.e. intrinsic parameters of each camera including the thermal one are known and that the relative positions between all cameras in the system are known [8]. Under the above conditions a 3D reconstruction of a thermal scene is straightforward: 3D coordinates of an object reconstructed by means of the stereoscopic set are remapped onto the image from thermal camera. The systems consisting of a pair of visual cameras and a single thermal camera are quite well documented [4,5,6,7,8]. On the other hand some authors [2] propose another approach i.e. stereoscopic set consisting of two thermal cameras. The main disadvantages of such systems are their complexity and the requirement of a specially designed tripod on which all required cameras should be installed.

In this article a novel approach, based on an integrated set of a single thermal and a single visual cameras is presented. Instead of a stereoscopic set a single camera is used for acquisition of two images of the same object from two different viewpoints. Depending on the observation conditions the system can use a pair of visual or thermal images to perform the 3D reconstruction.

2.   3D reconstruction by means of stereoscopic set of cameras

3D reconstruction of an observed scene by means of a stereoscopic set of two cameras is well known and documented [1,2,3]. A point P of an object can be localized in 3D space if two following conditions are met: (1) a point P is visualized in both cameras; (2) the cameras are calibrated i.e. their intrinsic and extrinsic parameters are known. Extrinsic parameters of a stereoscopic set are defined by 3x3 rotation matrix R and 3-dimensional translation vector T describing relative position of the two cameras. Intrinsic parameters describe the relation between 3D metric coordinates of visualized point and its respective image coordinates given as pixel position on camera's detector array [1,2,9].

The simple pin-hole model of camera is presented in figure 1a. The figure 1b shows the relation between metric coordinate system x, y of the image plane and the pixel array x_p, y_p of the camera sensor. The intrinsic parameters are grouped in the camera matrix relating pixels coordinates $p(x_p,y_p)$ to 3D space coordinates of visualised points $P(X,Y,Z)$ in the following way:

$$Z\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{1}$$

Where $f_x = f / h_x$, f – focal length of the camera lens; h_x – horizontal size of the pixel, $f_y = f/h_y$, h_y – vertical size of the pixel, c_x, c_y coordinates of the center of the image plane.

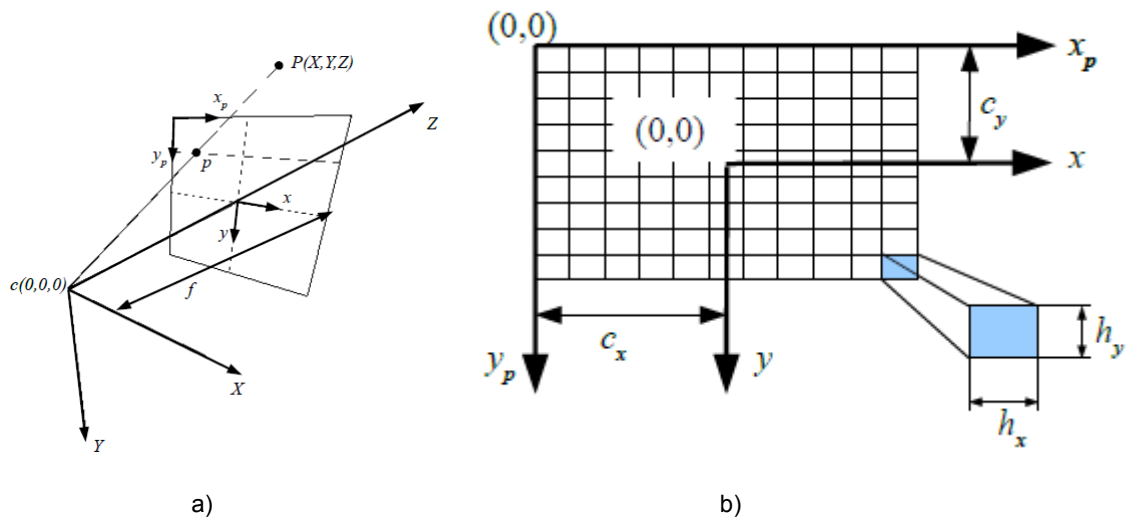a)   b)

Fig. 1 *a) Pin-hole model of a camera. b) Relation between the metric coordinate system of image plane of the camera and the pixel array of the camera sensor*

Figure 2 shows the principles of 3D reconstruction by means of a stereoscopic set of cameras. The point P_w is projected on the two image planes located in two different positions related by means of rotation matrix R and translation vector T. The relation between points in the coordinate systems of the first and the second cameras is as follows:

$$P = RP' + T \qquad (2)$$

Where $P = [X,Y,Z]^T$, $P' = [X',Y',Z']^T$ are 3D point coordinates with respect to first and second camera coordinate systems. 3D reconstruction refers in this case to evaluating either P or P' knowing the image coordinates p, p' in both views, relative position defined by R, T and the cameras matrices (defined by Eq. (1)) or one single camera matrix depending on whether the stereoscopic set consists of two different cameras or one single camera in two different positions.
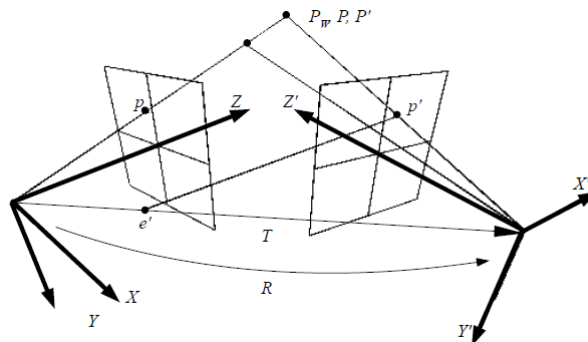

Fig. 2. *Stereoscopic set of cameras. Point P_w is visualised in two image planes located in two different positions related by means of rotation matrix R and translation vector T*

3.   Description of the system

In the presented approach it is assumed that the thermal and visual (if present) cameras are integrated and calibrated. Yet, the 3D reconstruction procedure is performed on image pairs given either by a thermal or a visual camera which are put in two different viewpoints (figure 3). The two thermal or visual images constitute stereoscopic pair which is only calibrated according to intrinsic parameters but not according to extrinsic ones because the matrix R and vector T of the same camera placed in two different positions are unknown.
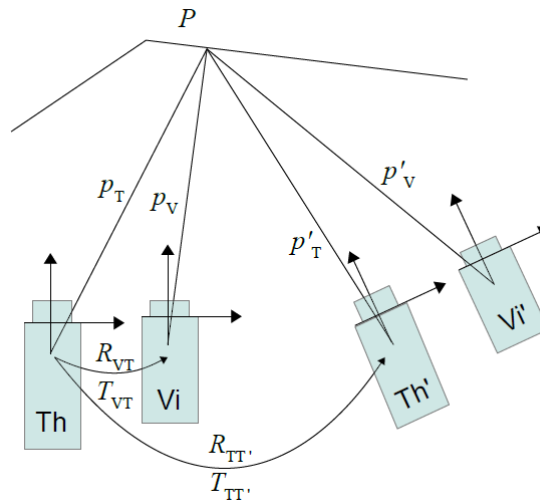
Fig. 3. *3D reconstruction with an integrated set of visual and thermal cameras located in two different positions. There are two stereoscopic sets: $R_{TT'}, T_{TT'}$ – consisting of thermal camera observing the reconstructed object from two different viewpoints; R_{VT}, T_{VT} – consisting of integrated visual and thermal cameras*

The reconstruction procedure consists of following steps:

1. Localization of feature points in thermal image and visual image for the first position of the set. It is worth notifying that depending on the scene, thermal image or the visual one can be more abundant with feature points.
2. The localized feature points are tracked in the respective images for the second position of the set and thus the correspondence between feature points for the two viewpoints is established.
3. The rotation matrix R and translation vector T are extracted either from homography matrix H or essential matrix E depending on the shape of the scene [2]. If the visualized feature points belong to the same plane in 3D space then the homography matrices are computed for each set of corresponding feature points: one homography H_T for thermal feature points' pairs, one H_V for the visual. If the visualized feature points are not from the same plane in 3D space then the essential matrices E_T and E_V are computed instead of the H matrices. For both sets of feature points pairs: visual and thermal and respective H_T, H_V or E_T, E_V matrices a reprojection error is computed. The outliers of corresponding feature points used for evaluation of respective matrices are removed by means of RANSAC method [2,9].
4. The initial matrix R and vector T are extracted from either E or H with the smallest reprojection error and then refined by gradient descent method. The vector T is extracted up to a scaling factor. In order to scale the vector T two methods are proposed (figure 4): (a) based on a known distance between two points in 3D space and (b) based on an additional approximate 3D reconstruction of a given point by the thermal and visual camera pair treated as a stereoscopic set.

4.   Scaling the translation vector

As it was mentioned earlier, in order to scale the translation vector T two methods are proposed. The drawback of the first method based on known distance between two points is that the image coordinates of the points must be known for both positions of the camera (figure 4a) – in the presented case the image coordinates were just marked manually by the user of the application.

In the second case (figure 4b) it is necessary to reconstruct at least one point by establishing its image coordinates both in visual and thermal images, as it is assumed that the thermal and visual cameras are calibrated against each other i.e. their relative position is known. On the other hand, the point can also be reconstructed basing on its image coordinates taken from both positions of the same camera (thermal or visual), but only up to an unknown scale factor. Comparing two reconstructed point 3D coordinates provided by means of the two stereoscopic sets (R_{VT}, T_{VT} and $R_{TT'}$, $T_{TT'}$) the scale factor can be evaluated.

The reconstruction of 3D points with a stereoscopic set consisting of visual and thermal cameras is not straightforward as not all the points of interest are visible in both images. To improve the reconstruction efficiency a novel method of establishing the correspondence between thermal and visual images is proposed. It is assumed that the pair consisting of visual and thermal images is rectified and therefore the corresponding points p_V, p_T representing some point P in 3D space must lay on the same horizontal lines in both rectified images. The method consists of the following steps:

1. From both images, thermal and visual, contours are extracted. A single contour is defined as polygon i.e. a sequence of vertices. Extracted contours from the thermal image are compared with the ones from the visual image.
2. The compared contours are assumed to be matched if their corresponding top and bottom boundaries are on the same horizontal lines.
3. From pairs of contours matched in this way, corresponding points are extracted. It is assumed that the corresponding points are located in the middle of the line segments connecting two vertices of contours if the difference of vertical components of vertices' coordinates is high enough.
4. For the matched contours from visual and thermal pair the corresponding contour and the corresponding point from that contour is found in the image from the camera located in the second position. If the evaluation of the camera displacement was done by means of feature points from the thermal images (as it is assumed in the figure 4b) then two correspondences for two stereoscopic sets are established: p_T, p_V and p_T, p'_T; the first correspondence provides 3D reconstructed point up to some scaling measure while the second one provides the 3D reconstructed point up to undefined scaling factor.
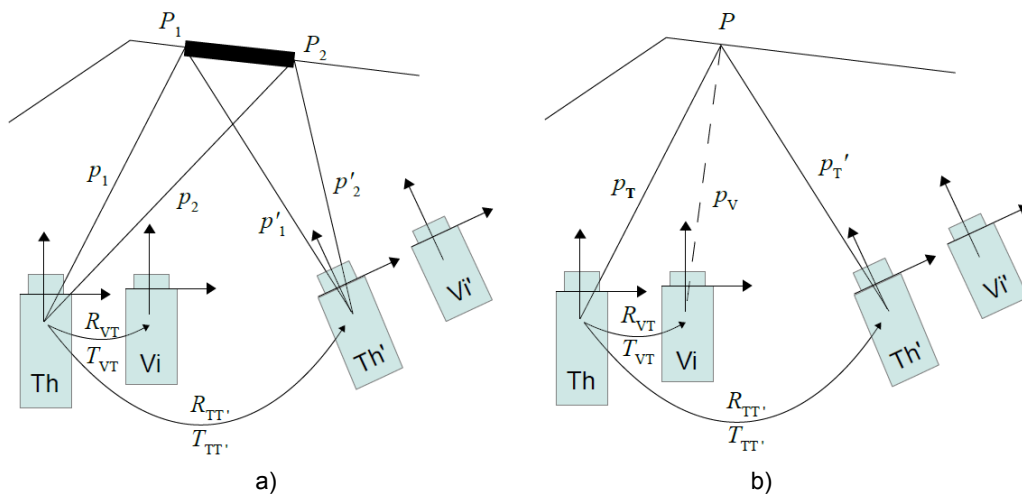

Fig. 4. *Scaling the translation vector T_{TT} : a) based on a known distance between two points P_1, P_2 in 3D space b) based on an additional approximate 3D reconstruction of a given point P by the thermal and visual camera pair treated as a stereoscopic set with R_{VT}, T_{VT} extrinsic parameters and p_T and p_V image coordinates.*

5.   Results

The proposed method was implemented in C++ using the OpenCV library [6]. The thermal camera used in the system was InfraTec VarioCam®, of 640x480 resolution. The used thermal camera is manufactured with integrated visual camera but it was not possible to get access to visual images by means of SDK delivered by the producer; therefore additional fixed camera, PointGrey Fly was used instead. Figure 5 shows the set of cameras mounted on special tripod used as a test bed of a 3D reconstruction concept presented in this paper.
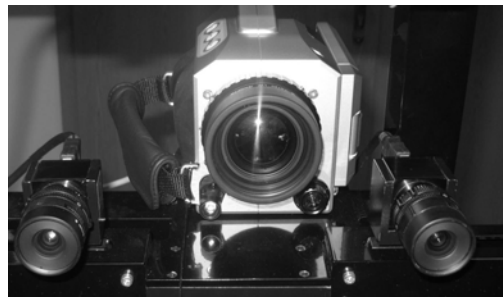

Fig. 5. *Three cameras (two visual one thermal) used as a test bed of the presented method of 3D reconstruction.*

In the system presented in figure 5 only the rightmost visual camera (the one fixed closer to the thermal camera) belongs to the visual-thermal stereoscopic set for evaluation of the aforementioned scaling factor. The second

visual camera is used to constitute classical stereoscopic set used for the verification of the results obtained by the proposed method.

The whole system of the cameras was calibrated using a set of algorithms provided by the OpenCV library [9]. The algorithms were implemented and integrated with specially designed graphical user interface by means of Builder C++ IDE. Figure 6 shows the user interface of the application during the calibration of a stereoscopic set consisting of visual and thermal cameras performed using a special calibration object. The calibration object has a typical chessboard [3,5,6,9] pattern with white squares covered with aluminum tape to lower the infrared emissivity to make the chessboard clearly visible in the thermal image. The temperature of the object was increased by the radiator put behind the board.
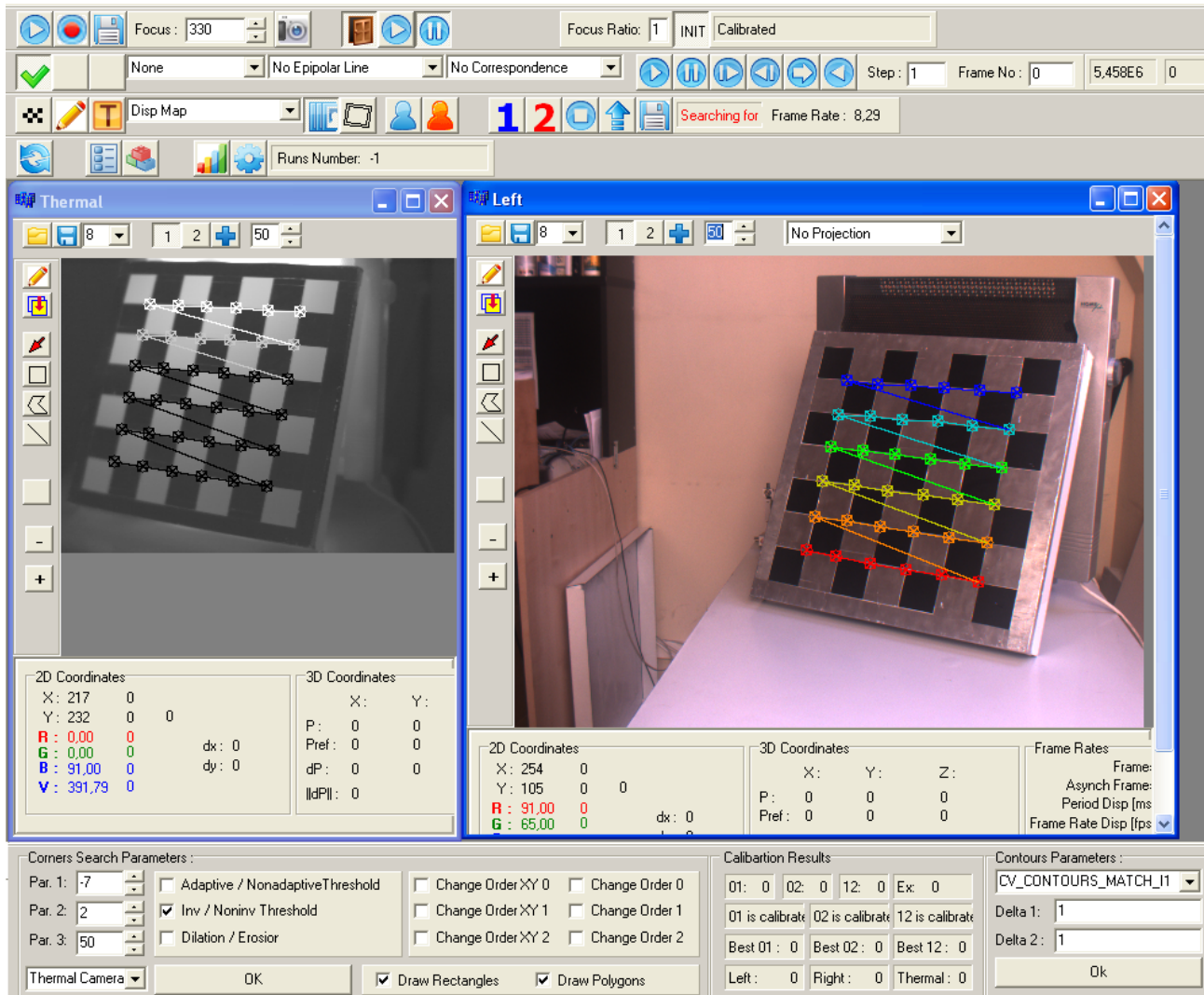

Fig. 6. *User Interface of the application designed for calibration, acquisition and 3D reconstruction.*

A calibration board was also used as a test object for acquisition of feature points with planar distribution and other non-planar objects were used to get feature points in 3D general positions. In figure 7a and 7b the images from thermal camera placed in two different positions are presented. The pair of the images a), b) is rectified [9,10] (the corresponding points are on the same horizontal lines) and is used for 3D reconstruction. The small white crosses indicate the feature points found in the original position of a camera (fig. 1a) and tracked with the camera moved to another position (fig.1b). The indicated feature points were found on the images before rectification as the rectification process is the result of evaluation of the camera displacement. Rotation matrix R and translation vector T describing the change of camera position were extracted from homography matrix H estimated by means of equation set derived from corresponding feature points coordinates. Refined values of the translation vector T and rotation vector describing rotation axis derived from R are presented in table 1. The extrinsic parameters after refinement with gradient descent method were evaluated with reprojection error e=0.047.
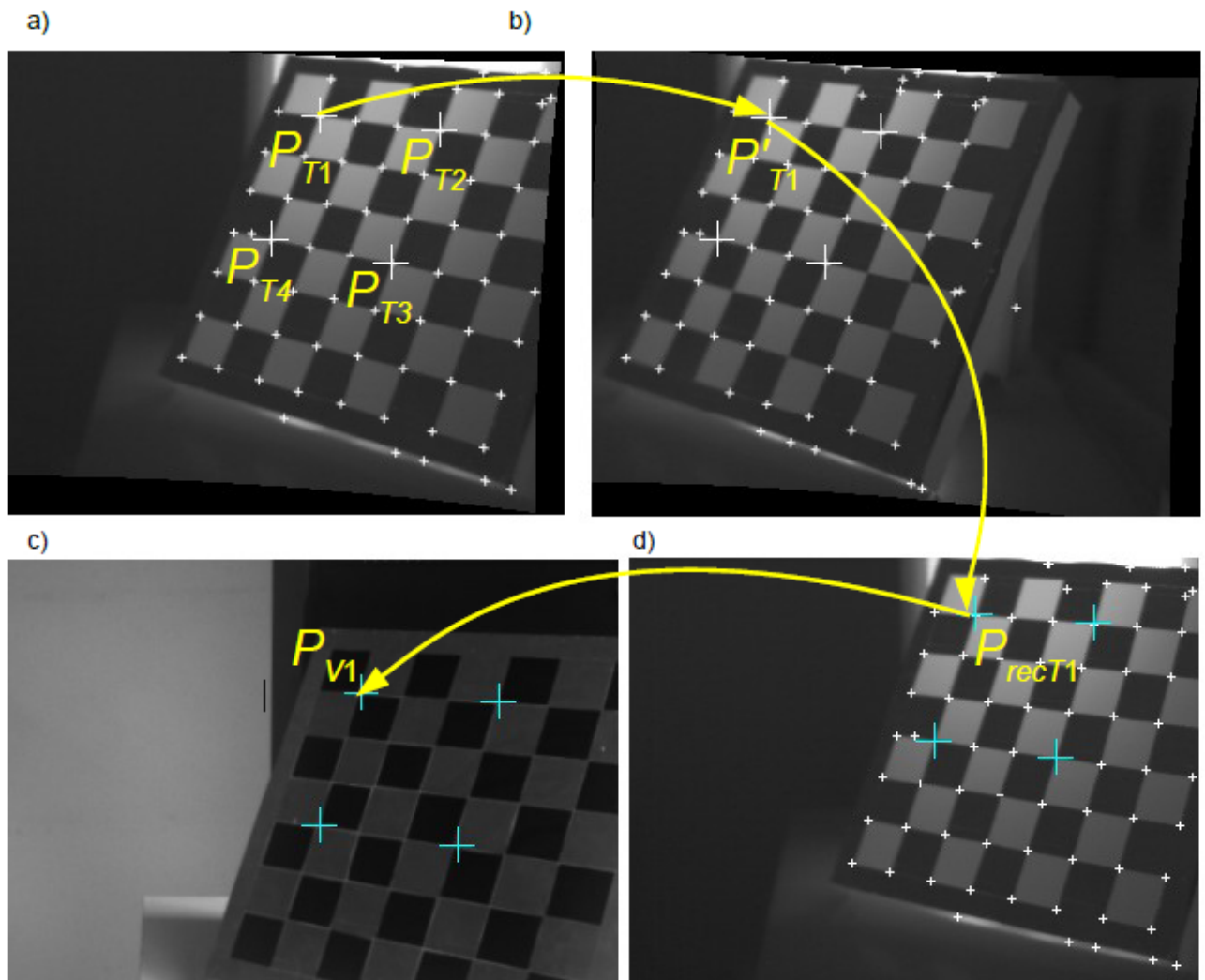
Fig. 7. *a) rectified thermal image given by camera in the first position, b) rectified thermal image given by camera in the second position, c) visual image given by camera in the first position d) thermal image given by camera in the first position before rectification*

Table 1. *Extrinsic parameters describing change of camera position*

Translation Vector	[183.6, -10.8, 7.6]
Rotation Vector	[-0.212, 0.946, 0.244]
Length of rotation vector expressing the angle of rotation [°]	4,9
Reprojection error	0,047

To evaluate the accuracy of the 3D reconstruction by means of the proposed method four characteristic points were reconstructed PT1, PT2, PT3, PT4. Figures 7a and 7b show the reconstructed points on the rectified pair of the thermal images; figure 7c shows the reconstructed points re-projected on the thermal images of the camera in the first position before rectification and the figure 7d shows the reconstructed points re-projected on the visual image of the camera constituting stereoscopic set with the thermal one. Tables 2 and 3 present the results of the reconstruction of the four points by means of three different stereoscopic sets: a) stereoscopic set provided by two rectified images from thermal camera in two different positions, b) stereoscopic set consisting of integrated visual and thermal cameras and c)

classical stereoscopic set consisting of two visual cameras. Table 2 compares the 3D coordinates of the reconstructed points for the stereoscopic sets while the table 3 compares the distances between the reconstructed points. As the real distance between all four points was the same and equal 180mm, comparing the real value with the one from the table 3, the quality of reconstruction can be assessed. Table 4 compares the lengths of the bases of the three stereoscopic sets, where the base of the stereoscopic set is defined as the length of the translation vector between the cameras. In the case of a single thermal camera operating from two different viewpoints the base was scaled by means of the single reconstructed point by thermal-visual stereoscopic set.

Table 2. *3D coordinates of reconstructed points X,Y,Z [mm]*

Point no	Single thermal camera	Thermal-visual set	Visual-visual set
PT1	[37.2, -192.1, 1520,1]	[37.5, -193.4, 1541.1]	[34.8, -184.8, 1487.9]
PT2	[187.2, -169.3, 1416.4]	[193.1, -170.6, 1455.2]	[182.9, -163.5, 1392.4]
PT3	[130.4, -0.5, 1348.3]	[132.5, 2.9, 1366.7]	[131.4, 2.1, 1358.4]
PT4	[-17.5, -21.7, 1478.4]	[-17.7, -19.2, 1480.2]	[-19.6, -19.0, 1449,3]

Table 3. *Distances between the reconstructed points [mm]*

Distance	Single thermal camera	Thermal-visual set	Visual-visual set
PT1, PT2	183.8	179.2	177.4
PT2, PT3	190.6	204.0	177.6
PT3, PT4	198.2	189.5	177.5
PT4, PT1	183.8	192.6	178.8

Table 4. *Comparison of the lengths of translation vectors of stereoscopic sets [mm]*

Single thermal camera	Thermal-visual set	Visual-visual set
184.1	90.1	195.9

Analysing the table 3 it can be concluded that the classical stereovision set gives the best reconstruction results – all distances are almost equal end are slightly different from the real value – the difference is about 3mm. Such system requires however three cameras to operate in the infrared (two visual and one thermal) and the reconstruction based only on visual images is not always valid for the thermal scene. The reconstruction with a single thermal camera operating from two different viewpoints is less precise but its direct operation on thermal images makes it independent of the scene conditions. The worst accuracy of the thermal-visual set result from the smallest base (shortest distance between the cameras) comparing to the other two stereoscopic sets (table 4) and therefore this reconstruction technique should be used only for approximate 3D reconstruction for scaling purposes.

6.   Conclusions

A system for 3D reconstruction of a thermal scene by means of a thermal-visual system of single cameras was presented. 3D scene is reconstructed by means of its feature points image coordinates delivered by the same thermal or visual camera observing the scene from two different viewpoints. The alternate use of visual and thermal cameras allows better recognition of feature points depending on the 3D scene conditions. The obtained results were compared with classical stereovision approach. The initial estimation of reconstruction error is up to 10%, while the important advantage of the proposed method is the use of only one thermal camera, which greatly reduces the cost of the system.

REFERENCES

[1] Cyganek B., Siebert J. P.,: "An introduction to 3D computer vision techniques and algorithms", John Wiley & Sons, Ltd, 2009.

[2] Ma, Y., Soatto, S., Kosecka, J., Sastry, S.: "An invitation to 3D vision, from images to models", Springer Verlag, 2003.

[3] Rzeszotarski D., Strumiłło P., Pełczyński P., Więcek B., Lorenc A.: „System obrazowania stereoskopowego sekwencji scen trójwymiarowych", Zeszyty Naukowe Elektronika, Zeszyt nr 10/ 2005, 2006, str. 165–181.

[4] Stein R., Sawicki P., Więcek B. "Directional Emissivity Correction by Photogrammetric 3D Object Reconstruction", Proc. Quantitative Infrared Thermography, QIRT'98, Eurotherm Seminar No 60, Łódź, Poland, Międzynarodowy Komitet Naukowy 7-10 września, 1999.

[5] Rzeszotarski D., Więcek B.: „Trójwymiarowa rekonstrukcja obrazów termowizyjnych z wykorzystaniem systemu stereowizyjnego", VII Konferencja Krajowa Termografia i termometria w podczerwieni 16-18 listopada 2006, str. 229-234.

[6] Rzeszotarski D., Więcek B.: "Thermal Images 3D reconstruction with use of stereovision system", 15th International Conference on Thermal Engineering and Thermogrammetry (THERMO), pp. 27–29 June 2007, Budapest.

[7] Prakash S., Lee Y., Robles-Kelly A.: "Stereo techniques for 3D mapping of object surface temperatures", QIRT Journal. Volume 4 – N° 1/2007, pp. 63–84.

[8] Rzeszotarski D., Więcek B.: "Calibration for 3D reconstruction of thermal images", QIRT 2008. 9th International Conference on Quantitative InfraRed Thermography.

[9] Bradski G., Kaehler A.: „Learning OpenCV", O'Reilly Media, 2008.

[10] Fusiello, A., Trucco, E. and Verri, A. : A compact algorithm for rectification of stereo pairs, Machine Vision and Applications 2002, vol. 12, pp. 16–22.