

Single thermal and visual camera system for 3D image fusion

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

This paper presents a novel approach based on the idea of 3D scene reconstruction with use of a system consisting of a single thermographic camera integrated with a single visual camera. Each camera acquires a sequence of images of the same object while the system is being moved. Using two separate Extended Kalman Filters, displacements of the visual and thermal camera relative to initial position can be estimated up to rotation and direction. A method of recovering metric scale by means of image coordinates of feature points tracked independently in two sequences is introduced. An example of thermal and visual image fusion by means of 3D data obtained by the proposed method is presented.

1. Introduction

Thermal image 3D reconstruction systems find increasing variety of applications. They enable to estimate more accurately the temperature distribution on the object being examined (by taking into account the directional emissivity of the material) [21], or to locate more accurately the source of infrared radiation [2,3]. Existing systems consist of visual stereoscopic set coupled with thermographic camera. In case of such systems 3D coordinates of an object reconstructed by means of the stereoscopic set are remapped onto the image from thermal camera [1,3,13,14,16,17]. On the other hand some authors propose another approach i.e. stereoscopic set consisting of two thermographic cameras [12,14]. 3D reconstruction is also possible by means of a system comprising a single camera provided that camera position relative to the reconstructed object is changed at the time of image capture. However, in this case, the three-dimensional reconstruction can only be made up to an unknown scaling factor [2,5,6,10,11,15,20,22].

This paper presents a novel approach based on the idea of 3D scene reconstruction introduced earlier by the author [15]. The approach so far presented has been based on a system consisting of a single thermographic camera integrated with a single visual camera. Each camera acquires two images of the same object from two different viewpoints. The drawback of presented method is that in order to recover metric dimension of the observed scene it either requires additional metric information about the observed scene, or additional 3D reconstruction by means of both visual and thermal image [15]. The system presented in this paper includes an improvement that addresses this issue. An example of thermal and visual image fusion by means of 3D data obtained by the proposed method is presented.

2. Description of the system

The system proposed in this work consists of integrated pair of cameras: thermographic and visual (figure 1). The system combines the two concepts of 3D reconstruction: (1) the concept of reconstruction with the help of a moving camera, and (2) the concept of reconstruction using the traditional stereovision system replaced by a pair of cameras: visual and thermographic. In addition to 3D scene reconstruction the system enables to merge thermograms with visual images. It is assumed that the integrated set of two cameras is moving and thus the reconstruction of the observed scene can be made independently for two image sequences: visual and thermographic. Basing on the image data from visual and thermographic cameras the motion of two cameras can be estimated up to the rotation and direction of translation [10]. As it has been proven below, in a such case it is possible to calculate the length of translation in reference to the known and constant distance between the cameras. This means that knowing the constant distance between two integrated cameras, one can estimate the length of the path traveled by the system, which in turn means the possibility to merge the visual and thermal images.

During initialization, the system is set in initial position. From images acquired from respective cameras in initial position feature points are selected. Feature points are points that are unique e.g. corners and therefore can be tracked in the sequence of images [9,18].

Using two separate Extended Kalman Filters (EKF) [5,6,10], displacements of the visual and thermal camera relative to initial position can be estimated up to rotation and translation with given arbitrary scale. Spatial coordinates of tracked feature points are also estimated using the same EKFs. At this stage, the scale of the reconstructed points remains unknown, i.e. depth components of tracked feature points and length of the displacement vectors of two cameras are calculated relative to the fixed depth component of two arbitrarily selected feature points: $Z_{T1}(t) = Z_{init}$ (for visual camera) $Z_{T1}(t) = Z_{init}$ (for thermographic camera). These reference feature points are excluded from respective state vectors to obtain observability of the reconstruction system [5,6,10,19].

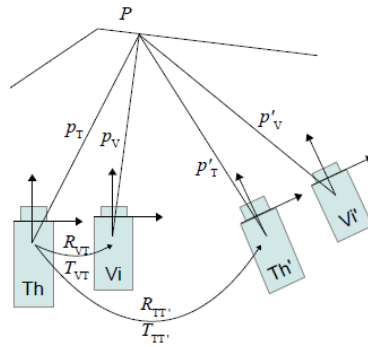


Fig. 1. 3D reconstruction by means of moving set of two cameras: thermographic coupled with visual

The translations with given arbitrary scales and matrices of rotation of both cameras estimated by means of EKF, together with the extrinsic parameters describing relative position of the cameras, enable to calculate the scales of the translation vectors and therefore their metric lengths. Finally, using estimated scales, the metric coordinates of feature points tracked in the images from thermal and visual cameras can be determined.

It is worth to emphasize that to estimate motion parameters of the camera being displaced apart of EKF, a method using an essential matrix can be used [10]. But as it is shown below, only motion parameters of the cameras estimated by means of EKFs were suitable for scale coefficients recovery.

3. Implementation details

In order to estimate the cameras displacements and positions of feature points being tracked independently in visual and thermal image sequences, two extended Kalman filters were implemented based on description provided by Chiuso et al [5,6,10]. Because of relation between 3D coordinates $\mathbf{X}_i=[X_i, Y_i, Z_i]^T$ of a given point P_i and respective normalized images coordinates $\mathbf{x}_i=[x_i, y_i]^T$ is as follows:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} \frac{X_i}{Z_i} \\ \frac{Y_i}{Z_i} \end{bmatrix}, \quad (1)$$

therefore it can be assumed that the 3D position of point P_i can be described by normalized image coordinate $\mathbf{x}_i=[x_i, y_i]^T$ and the depth component Z_i of 3D coordinate [10]. In order to obtain normalized coordinate of a feature point being tracked in the image sequence provided by the camera, the camera must be calibrated i.e. its intrinsic parameters must be estimated [7,10]. Taking into account the issues mentioned above two state vectors $\chi_V(t), \chi_T(t)$ at time step t for visual camera and thermal camera respectively can be defined as follows:

$$\chi_V(t) = [x_{V1}^T(t), \dots, x_{VN}^T(t), Z_{V1}(t), \dots, Z_{VN}(t), T_V^T(t), \Phi_V^T(t), v_V^T(t), \omega_V^T(t)]^T, \quad (2)$$

$$\chi_T(t) = [x_{T1}^T(t), \dots, x_{TM}^T(t), Z_{T1}(t), \dots, Z_{TM}(t), T_T^T(t), \Phi_T^T(t), v_T^T(t), \omega_T^T(t)]^T, \quad (3)$$

where N, M are numbers of feature points being tracked independently in image sequences provided by visual and thermal camera respectively; $\mathbf{x}_{Vi}(t)=[x_{Vi}(t), y_{Vi}(t)]^T, \mathbf{x}_{Ti}(t)=[x_{Ti}(t), y_{Ti}(t)]^T$ are normalized image coordinates of feature points in images provided by visual camera and thermal camera respectively being in the initial position, i.e. at time step $t=0$; $Z_{Vi}(t), Z_{Ti}(t)$ are depth components of 3D coordinates of feature points tracked in visual and thermal sequence respectively in reference to initial position of respective camera. $T_V(t)=[T_{Vx}(t), T_{Vy}(t), T_{Vz}(t)]^T, \Phi_V(t)=[\Phi_{Vx}(t), \Phi_{Vy}(t), \Phi_{Vz}(t)]^T$ are motion parameters of visual camera i.e. translation and rotation vectors describing current position (at time step t) of the camera against the initial position (at time step $t=0$). Similarly $T_T(t), \Phi_T(t)$ are motion parameters of thermal camera. $v_V(t), \omega_V(t)$ and $v_T(t), \omega_T(t)$ are vectors of linear and angular velocities by means of which a relation between two adjacent positions (e.g. at time steps t and $t-1$) of the given camera can be expressed. Therefore a relation between 3D coordinates $\mathbf{X}'_{Vi}(t), \mathbf{X}'_{Ti}(t)$ in reference frame associated with

camera being in time step t and 3D coordinates $X_{Vi}(t)=X'_{Vi}(0)$, $X_{Ti}(t)=X'_{Ti}(0)$ in reference frame associated with camera in initial position can be expressed as follows for visual and thermal camera respectively:

$$X'_{Vi}(t)=R_{Vi}(t)X_{Vi}(t)+T_{Vi}(t), \quad (4)$$

$$X'_{Ti}(t)=R_{Ti}(t)X_{Ti}(t)+T_{Ti}(t), \quad (5)$$

where $R_{Vi}(t)$, $R_{Ti}(t)$ are rotation matrices corresponding to rotation vectors $\Phi_{Vi}(t)$, $\Phi_{Ti}(t)$ by means of Rodrigues' formula.

The measurement (observation) vectors for each camera at given time step consist of normalized image coordinates of feature points being tracked in images provided by cameras being in corresponding positions:

$$\psi_V(t)=\left[x'_{V1}(t), \dots, x'_{VN}(t)\right]^T \quad (6)$$

$$\psi_T(t)=\left[x'_{T1}(t), \dots, x'_{TN}(t)\right]^T \quad (7)$$

Summing up there are two independent 3D reconstruction systems for each camera (visual and thermal). Each system is modeled by a pair of state and observation vectors (Eqs. (2) and (6) or Eqs. (3) and (7)). It can be proven [5,6,10] that the system modeled with use of so defined state vector and observation vector is not observable which means that state vector of the system can not be determined from the observation vector of the system. It also has been proven [5,6,10] that the analyzed system will be observable if three image coordinates and one depth component of arbitrary chosen feature points are excluded from the state vector. Thus the state vectors of the two observable systems for both cameras can be defined as follows:

$$\chi_V(t)=\left[x_{V4}^T(t), \dots, x_{VN}^T(t), z_{V2}(t), \dots, z_{VN}(t), T_V^T(t), \Phi_V^T(t), v_V^T(t), \omega_V^T(t)\right]^T, \quad (8)$$

$$\chi_T(t)=\left[x_{T4}^T(t), \dots, x_{TN}^T(t), z_{T2}(t), \dots, z_{TN}(t), T_T^T(t), \Phi_T^T(t), v_T^T(t), \omega_T^T(t)\right]^T, \quad (9)$$

where it has been assumed that indexes of excluded feature points equal 1,2, and 3.

Using state and observation vectors defined for both cameras according to Eqs. (6) to (9) along with proper transition/observation models [5,6,10] and extended Kalman filter predict/update equations [5,6,10] the 3D positions of feature points tracked by each camera and their motion parameters can be determined up to unknown scale. The initial state vectors for each camera were chosen as follows:

$$\chi_V(0)=\left[x_{V4}^T(0), \dots, x_{VN}^T(0), z_{init1}, \dots, z_{init1}, 0,0,0,0\right]^T, \quad (10)$$

$$\chi_T(0)=\left[x_{T4}^T(0), \dots, x_{TN}^T(0), z_{init1}, \dots, z_{init1}, 0,0,0,0\right]^T, \quad (11)$$

where initial image coordinates are the image coordinates of feature points detected in images from cameras being in position at initial time step $t=0$. The initial values of depth components equal the arbitrary chosen value of depth components $Z_{V1}(t)=Z_{init1}$, $Z_{T1}(t)=Z_{init1}$ excluded from the respective state vectors in order to obtain observability of the two systems. In practice that means that the depth components of feature points and the length of current displacement of each camera are estimated relative to arbitrary chosen reference depth Z_{init1} . In order to merge together estimated 3D coordinates of feature points tracked in the two image sequences (visual and thermal) a true scale has to be calculated.

A novel method of true 3D scene scale recovery with use of motion parameters of the two integrated cameras is introduced below. The system of the two cameras is presumed to be calibrated, i.e. its intrinsic (optical) parameters and extrinsic parameters (relative positions of the cameras) are known [7,10,16]. The system is calibrated and therefore at a time step t a relation between 3D coordinates $X'_V(t)$ in reference frame of the visual camera and 3D coordinates $X'_T(t)$ in reference frame of the thermal camera can be simplified to a so called canonical form [7]:

$$X'_V(t)=X'_T(t)+T_{VT}, \quad (12)$$

where T_{VT} is a translation vector in a canonical stereoscopic [7] set of visual and thermal camera, $T_{VT}=[\|T_{VT}\|, 0, 0]^T$. A relation between 3D coordinates $X'_V(0)$ in reference frame of visual camera being in initial position (at time step $t=0$) and 3D coordinates $X'_V(t)$ in reference frame of the same camera being displaced at time step t is as follows:

$$X'_v(0) = R_v(t)X'_v(t) + \alpha_v(t)T_{V(Z_1)}(t), \tag{13}$$

where $R_v(t)$, $T_{V(Z_1)}(t)$ are motion parameters of visual camera determined with use of EKF; translation vector $T_{V(Z_1)}(t)$ is estimated in relation to arbitrary chosen reference depth component $Z_{v1} = Z_{init1}$. $\alpha_v(t)$ is a scale factor relating estimated translation vector $T_{V(Z_1)}(t)$ to the vector $T_v(t)$ associated with unknown true dimensions of the scene being reconstructed. As long as the reference feature point with depth Z_{v1} is the same, the scale factor remains unchanged for each time step during displacement of the system i.e.: $\alpha_v(t) = \text{const}$.

Similar relation can be found for thermal camera:

$$X'_t(0) = R_t(t)X'_t(t) + \alpha_t(t)T_{T(Z_1)}(t), \tag{14}$$

Substituting Eq. (12) into Eq. (13) a relation between motion parameters of thermal camera and motion parameters of visual camera can be found as:

$$R_v(t) = R_t(t), \tag{15}$$

$$\alpha_t(t)T_{T(Z_1)}(t) = R_v(t)T_{VT} - T_{VT} + \alpha_v(t)T_{V(Z_1)}(t), \tag{16}$$

Eq. (16) can be transformed into:

$$\begin{bmatrix} T_{T(Z_1)}(t) & -T_{V(Z_1)}(t) \end{bmatrix} \begin{bmatrix} \alpha_T(t) \\ \alpha_V(t) \end{bmatrix} = \|T_{VT}\| \begin{bmatrix} r_{V11}(t) - 1 \\ r_{V21}(t) \\ r_{V31}(t) \end{bmatrix}, \tag{17}$$

where $r_{v11}(t)$, $r_{v21}(t)$, $r_{v31}(t)$ are elements of rotation matrix $R_v(t)$. Eq. (17) is in fact a set of three equations with two unknowns $\alpha_v(t)$, $\alpha_t(t)$ and can be solved using least square method. Eq (17) shows that it is possible to recover true scale of the scene using integrated set of two cameras without finding correspondence between feature points detected in pair of images providing that the set is calibrated i.e. relative positions of the cameras is known including the distance $\|T_{VT}\|$. This feature is especially important in case of integrated set of visual and thermal cameras, as establishing correspondences between pairs of feature points detected in visual image and thermal image is very difficult [15].

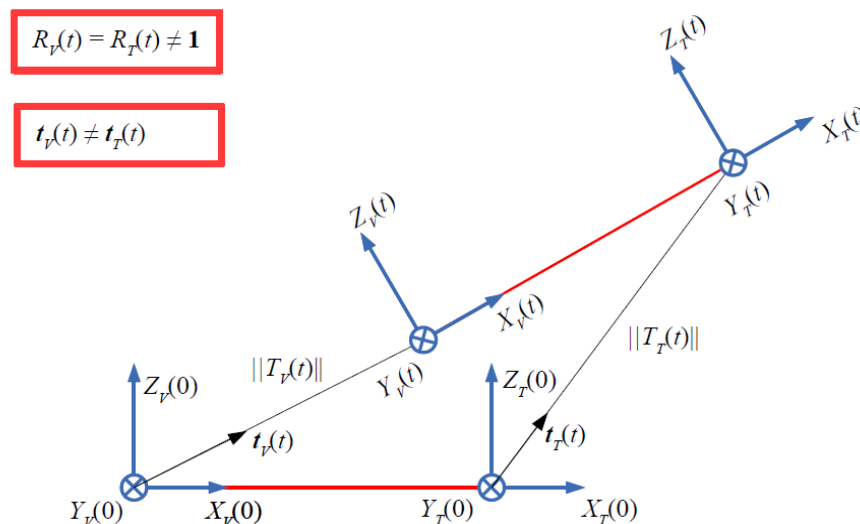


Fig. 2. Displacement of integrated set of two cameras with nonzero rotation and different directions of translation vectors

Figure 2 shows graphical interpretation of Eq. (17). Two positions of the integrated set of two cameras are illustrated with use of reference frames' axis associated with the cameras: $X_v(0)$, $Y_v(0)$, $Z_v(0)$ – visual camera at time step $t=0$, $X_t(0)$, $Y_t(0)$, $Z_t(0)$ – thermal camera at time step $t=0$, $X_v(t)$, $Y_v(t)$, $Z_v(t)$ – displaced visual camera at time step $t>0$, $X_t(t)$, $Y_t(t)$, $Z_t(t)$ – displaced thermal camera at time step $t>0$.

Eq. (17) has a single nonzero solution if the directions $\mathbf{t}_V(t)$, $\mathbf{t}_T(t)$ of translation vectors $T_{V(Z1)}(t)$, $T_{T(Z1)}(t)$ are different and there is nonzero rotation: $R_V(t)=R_T(t)\neq\mathbf{1}$. If the translation vectors are parallel then Eq. (17) has infinite number of solutions.

4. Results

A testbed system consisting of two cameras (one visual, one thermographic) has been built (figure 3). The thermal camera used in the system was InfraTec VarioCam, of 640x480 resolution. The visual camera used in the system was PointGrey Fly, of resolution 1024x768 resolution. An important part of the system is a software application designed with use of OpenCV library, in order to perform calibration of the set of cameras, capture and record image sequences and to estimate motion parameters and 3D positions of feature points being tracked independently in image sequences by means of extended Kalman filter and essential matrix.

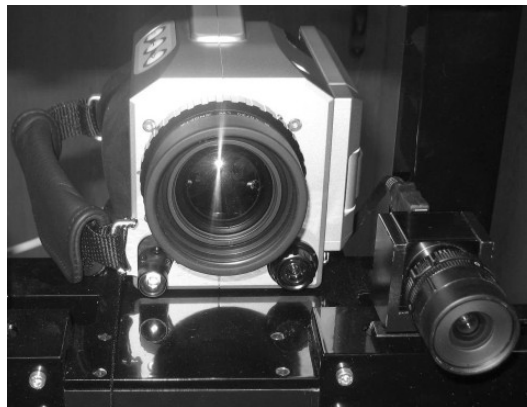
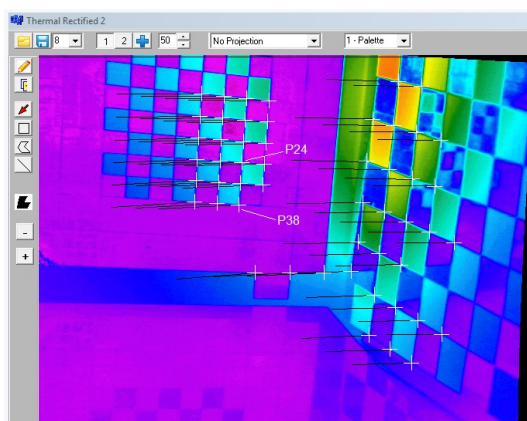
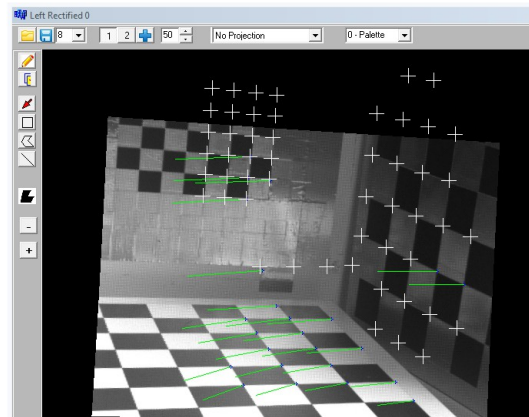


Fig. 3. A test system consisting of thermal and visual cameras

In order to verify the proposed method a test scene has been built comprising three planes composed of calibration boards (figure 4). Feature points on two vertical boards are visible both in visual image and thermographic image due to different emissivity of dark and bright squares – bright squares were covered with aluminum foil and additionally the boards were heated with radiator. On the other hand feature points on horizontal board are only visible in visual image and therefore this plane can be reconstructed with use of visual data only. But due to true scale recovery by the proposed method reconstructed points can be reprojected on to the thermal image.



a)



b)

Fig. 4. A test scene consisting of three calibration boards. a) thermal image, b) visual image)

Figures 5, 6 show an example of estimation of distance between two points marked on figure 4 as P24, P38. The distance scale was estimated at each time step t_N by means of the proposed method with use of motion parameters

determined either by essential matrix (figure 5) or extended Kalman filter (figure 6). In either case (EKF or EM) the results are compared with the results obtained when the scale coefficients were presumed to be known.

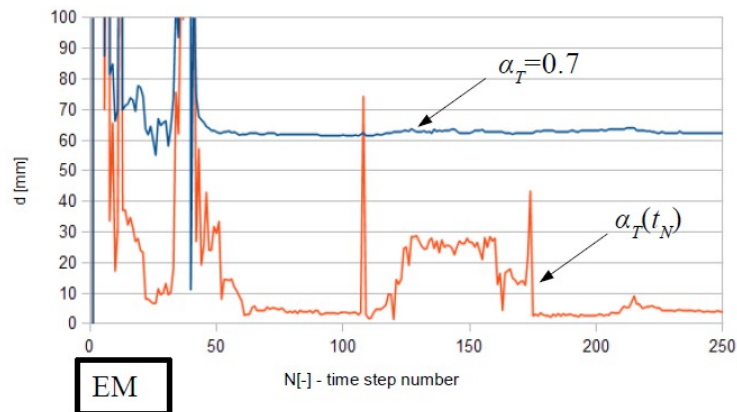


Fig. 5. Distance between two points reconstructed with use of essential matrix, $\alpha_T=0.7$ – scale coefficient known, $\alpha_T(t_N)$ – scale coefficient calculated at each time step by means of proposed method

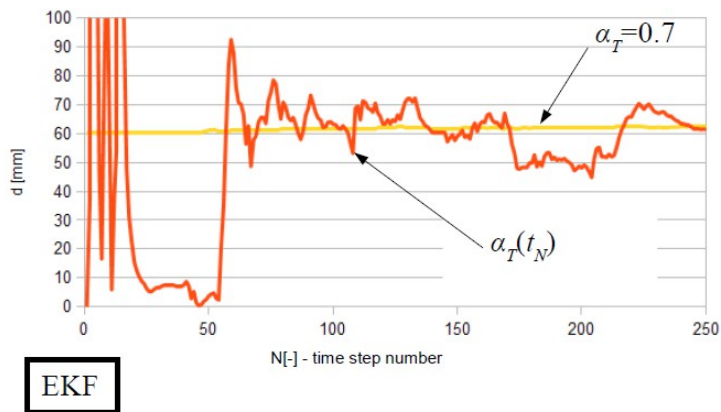


Fig. 6. Distance between two points reconstructed with use of extended Kalman filter, $\alpha_T=0.7$ – scale coefficient known, $\alpha_T(t_N)$ – scale coefficient calculated at each time step by means of proposed method

Because the real distance between test points was known (60 mm), an average relative error of distance obtained by either method could be calculated and is summed up in tables 1 and 2. The average error was calculated for such a range of time step numbers N that extended Kalman filter operation was stable i.e. bereft of large errors that occurred at the beginning ($N < 70$).

Table 1. Relative average error of distance between two points reconstructed when scale coefficients were known, $\alpha_T=0.7$

Method	EM	EKF
$\Delta d/d$ [%]	4.72	3.44

Table 2. Relative average error of distance between two points reconstructed when scale coefficients were calculated at each time step by means of proposed method

Method	EM	EKF
$\Delta d/d$ [%]	80.5	8.39

The results show that when the scale coefficients are known the average relative errors are comparable for EKF and EM (table 1). On the other hand table 2 shows that the average relative errors are acceptable only when the motion parameters are obtained with use of EKF (table 2).

Figure 7 shows an example of application of 3D reconstruction of tracked feature points to merge selected regions of visual and thermographic image. In figure 4a) regions of visual image which belong to reconstructed planes S1, S2, S3 are selected. The equations of the planes S2 and S3 were estimated by means of feature points tracked in thermographic image while the equation of plane S1 was estimated by means of feature points tracked in visual image – it is worth to emphasize that only feature points visible in visual image were suitable for 3D reconstruction of plane S1.

Figure 4b) shows thermographic image merged with selected regions of visual one.

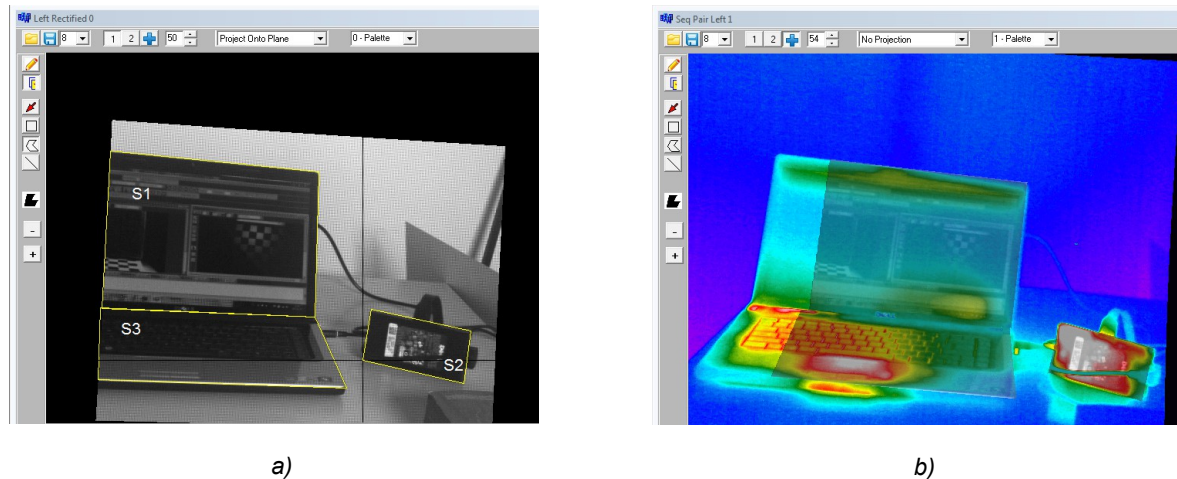


Fig. 7. An example of merging visual image with thermogram. a) Visual image in initial moment ($N=0$) with selected regions which are to be merged with thermogram. b) Thermogram captured in moment of time no $N=186$ merged with selected regions of visual image

5. Conclusions

A new approach of 3D thermal scene reconstruction with use of image sequences from thermographic and visual camera was presented. The system consists of integrated set of thremographic and visual camera being moved relative to reconstructed object.

The key element of the system is a new, original method to estimate a displacement of an integrated set of two cameras. The main advantage of the method is the ability to recover the metric dimensions of the scene being observed with the use of image coordinates of feature points tracked independently in images being captured by each camera. Thus the method may also be used in mobile visual system comprising cameras with different parameters or with non-overlapping fields of view.

REFERENCES

- [1] Aksenov P., Clark I., Grant D., Inman L.V.A., and Nebel J.C., 3d thermography for the quantification of heat generation resulting from inflammation, Proceedings of 8th 3D Modelling symposium, Paris, France, 2003.
- [2] Azarbayejani A., Horowitz B., Pentland A., Recursive Estimation of Structure and Motion using Relative Orientation Constraints, Conference on Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society.
- [3] Barone S., Paoli A., Razonale A., A biomedical application combining visible and thermal 3D imaging, XVIII Congreso Internactional de Ingenieria Grafica, Sitges, 2006.
- [4] Bradski G., Kaehler A.: Learning OpenCV, O'Reilly Media, 2008.
- [5] Chiuso A., Favaro P., Jin H., Soatto S., 3-D Motion and Structure from 2-D Motion Casually Integrated over Time: Implementation, Computer Vision — ECCV 2000, Lecture Notes in Computer Science Volume 1843, 2000, pp 734-750.
- [6] Chiuso A., Favaro P., Jin H., Soatto S., Structure from Motion Casually Integrated Over Time, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 4, April 2002.
- [7] Cyganek B., Siebert J. P.: An introduction to 3D computer vision techniques and algorithms, John Wiley & Sons, Ltd, 2009.
- [8] Davison A.J., Reid I.D., Molton N.D., Stasse O., MonoSLAM: Real-Time Single Camera SLAM, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 6, June 2007.
- [9] Harris C., Stephens M. A combined corner and edge detector, Proceedings of the 4th Alvey Vision Conference, 1988.
- [10] Ma Y., Soatto S., Kosecka J., Sastry S.: An invitation to 3D vision, from images to models, Springer Verlag, 2003

- [11] Meier W., von Stein H.-D., Estimation of object and sensor motion in infrared image sequences, IEEE International Conference, Image Processing, 1994. Proceedings. ICIP-94.
- [12] Prakash S., Lee Y., Caelli T., Raupach T., Robust Thermal Camera Calibration and 3D Mapping of Object Surface Temperatures, SPIE 2006.
- [13] Prakash S., Lee Y., Caelli T.: 3D Mapping of Surface Temperature Using Thermal Stereo, 9th International Conference on Control, Automation, Robotics and Vision, 2006. ICARCV '06.
- [14] 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.
- [15] Rzeszutarski D., Więcek B.: An integrated thermal and visual camera system for 3D reconstruction, QIRT 2012. 11th International Conference on Quantitative InfraRed Thermography.
- [16] Rzeszutarski D., Więcek B.: Calibration for 3D reconstruction of thermal images, QIRT 2008. 9th International Conference on Quantitative InfraRed Thermography.
- [17] Rzeszutarski 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.
- [18] [26] Shi J., Tomasi C., Good features to track, 9th IEEE Conference on Computer Vision and Pattern Recognition, 1994.
- [19] Soatto S., 3-d structure from visual motion: modeling, representation and observability, Automatica, 33:1287-1312, 1997.
- [20] Soatto S., Perona P., Temporal integration of visual information for the robust recovery of motion and structure, CNS Tech Report n. 24, 1992.
- [21] 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.
- [22] Yiu-Ming, Ng H., Du R., Acquisition of 3D Surface Temperature Distribution of a Car Body, 2005 IEEE International Conference on Information Acquisition.