Warping-based co-registration of thermal infrared images: study of factors influencing its applicability

Cardone D.[1], Pinti P.[1], Di Donato L.[1], Merla A.[1]

1. Infrared Imaging Lab – Centro ITAB - Institute for Advanced Biomedical Technologies and Department of Neuroscience, Imaging and Clinical Sciences, University of Chieti-Pescara, Italy.

Correspondence and requests for materials should be addressed to D.C. (email: d.cardone@unich.it, telephone: +39 08713556954, fax: +39 08713556922)

Abstract – A relevant issue for processing biomedical thermal imaging data is the availability of tools for objective and quantitative comparison of images across different conditions or subjects. To this goal, a solution can be offered by projecting the thermal distribution data onto a fictitious template to obtain a common reference for comparison across cases or subjects.

In this preliminary study, we tested the feasibility of applying a warping procedure on infrared thermal images. Fifteen thermal images of checkerboard were recorded at three different distances and five different angles in order to evaluate which factor mostly influences the warping accuracy.

The accuracy of three different warping transformation models (local weighted mean (LWM), polynomial, affine) was tested by comparing the positioning error between users' selected fiduciary points on each thermal image and their corresponding reference position assigned on the template image.

Fifteen users, divided into three groups upon on their experience in thermal imaging processing, participated in this study in order to evaluate the effect of experience in applying a warping procedure to the analysis of thermal infrared images.

The most relevant factor influencing the positioning and thermal errors is the acquisition distance, while the users' level of experience and the inclination angle do not seem to play the same importance. Comparing the three transformations, the LWM seems to be the best in terms of minimizing the two categories of errors. This preliminary work helps to understand the limits and the possibilities of applying warping techniques for objective, quantitative and automatic thermal image comparisons.

Keywords: thermal imaging; warping; registration; local weighted mean; feasibility;

1. Introduction

Thermal InfraRed (IR) imaging or thermography is a wide-spread imaging technique used to accurately evaluate the thermal distribution of a body without any contact between the sensors and the body itself. Thermal imaging devices, or thermal cameras, are able to capture the body infrared radiation and convert it into a thermal image. Modern thermal IR cameras allow counting on both very high spatial (up to 1280x1024 pixel), temporal recording resolution (full frame frequency rate up to 200 Hz) and high thermal sensitivity (up to 15 mK @ 30°C) in the spectral range [3-5 and 7-14] µm [1]. Thermal IR imaging is used in various fields of applications, from the mechanical to the electrical, as from the industrial to the buildings to the biomedical fields [18].

A very relevant issue for advanced processing of thermal imaging is the availability of tools for objective and quantitative comparisons of thermal images and data among conditions or samples. This need is particularly important in the biomedical field, especially with human subjects. To this goal, a co-registration approach should be followed, either to co-register the thermal images among them for further processing or comparison, and to project even a full dataset over a reference template to obtain a group temperature distribution for a given condition. Such a capability allows, in fact, to overcome the operator-dependent effects (potentially occurring for region of interest-based analysis) and opens up the way to effective group analysis, the same way they are currently performed in neuroimaging through the use of brain atlases (e.g., Talairach coordinate system of the human brain [2]).

In this paper, we evaluated the feasibility of pursuing the co-registration of thermal images by applying an image warping approach.

Image warping is an imaging processing technique based on the determination of a spatial transformation which maps all the pixels of one image onto the pixels of another image. It is a very useful tool in the image processing field as it allows the co-registration of two or more images [4; 5; 6], to align an image with a reference grid such as a map or a template [5], or to remove optical distortions introduced by the acquisition device [7; 8].

Image warping might be provided by the detection and matching of some relevant features such as corresponding sets of points between images, local measurement of correlation between images or edges overlapping [5; 9; 10]. The following operations consist in (i) the estimation of the transformation model (mapping function), (ii) image resampling and (iii) image transformation [10].

The transformation model is chosen depending on the specific application and on the particular set of images to warp. The mapping function could be either global or local: The former uses all the control points (CPs) for the estimation of a set of parameters valid for the whole image, and the latter considers the image as a composition of patches (typically triangles), determining a function parameter for each patch.

There are several works concerning the registration between visible and thermal images. In 1989, Toet et al. introduced a hierarchical image merging scheme based on a multi-resolution contrast decomposition (the ratio of a low-pass pyramid) [12].
Over the next years, affine transformation methods have been proposed to register the thermal image on its corresponding visible one. For instance, in 2006, Schaefer et al. used an affine transformation to register thermal frames and visible images to obtain a fusion image [11]. In 2007, Istenic et al. employed a multi-sensor registration in the Hough-parameter space for building images [13], while Kong et al. utilized multiscale fusion of thermal and visible images to obtain a reliable face recognition, independent from illumination conditions. The combination of both imaging techniques improved the recognition performances under a wide range of illumination change [15]. In 2009, Howell et al. made use of an affine transformation for the visible-infrared registration to assess the skin temperature and blood flow in childhood localized scleroderma [14]. In 2013, Cheng et al. elaborated a motion tracking system through a template-based algorithm. They estimated the motion parameter of the template image using an affine warping model; Lucas-Kanade algorithm was then applied to search for the optimized parameters of the warping function [17]. More recently, in 2014, in order to monitor mental stress in a contact-less way, Mohd et al. developed a new approach in the registration of facial thermal-visible by using a nostril mask and by adopting the Scale Invariant Feature Transform (SIFT) for point extraction and matching. The percentage of correct registration matching was 86% [16].
Finally, first attempts for obtaining a group atlas of temperature distribution for the whole human body were also performed by Ring and colleagues [3]. They developed a procedure based on the identification and selection of 87 regions of interest on 27 specific views of the body, following the Glamorgan protocol. They created a first database of thermal distribution on subjects between 18 and 70 years old.

In this work, we aimed at demonstrating the feasibility of registering one or more thermal images onto a common template image. To this goal, we warped a sample of thermal images of a chessboard acquired at different distances and inclination angles on a reference template (reference chessboard). The recorded images were not corrected for thermal reflections and uniformity to replicate standard real-life experimental conditions.

To establish the feasibility of the warping process, we compared a set of points, manually positioned by operators (OP points), with a reference standard point distribution (REF points), obtained from the thermal chessboard with a semi-automated procedure. We then determined the positioning and thermal errors between the blobs derived from OP and REF points.

We estimated the same categories of errors (positioning and thermal errors) between OP points and the back-projection of the reference template points (Back Projection point, BP), obtained from each of the three transformations, to compare the performances of the three different mapping warping functions (affine and polynomial, i.e. global transformations, and local weighted mean (LWM), i.e. local transformation).

The paper is structured as follows: In the Material and Methods section, we describe the experimental technique and procedure, focusing on the definition of a reference standard, used to estimate the positioning and thermal errors respect to the real experimental measurement obtained by users. Positioning and thermal errors are also evaluated respect to the specific warping transformation. The Results section presents the statistical analyses conducted on the above-mentioned errors, while the Discussion and Conclusion section discusses and establishes the feasibility of the present study, and suggest which mapping transform performs best in terms of positioning and thermal errors.

2. Materials and Methods

2.1 Procedure

A home-made thermal chessboard was used for the purpose of this study. It was built on a plexiglass plane surface, surrounded by a cardboard, and it was composed of 108 blobs (~ 20 mm^2 each; vertical side dV=20 mm, horizontal side dH=19 mm), realized by the alternation of two different materials with different emissivity: aluminum sheet on double-sided tape blobs and white painted plexiglass blobs (Fig. 1 (a)). The alternation of the two different materials allowed the clear distinction between blobs in the thermal images (Fig. 1 (b)). In order to always place the chessboard with the same orientation during the acquisition, a distinction element was included in the right bottom corner of the chessboard, consisting in 3 not-alternating blobs. The checkerboard was composed of 9 rows and 12 columns.

(a)   (b)
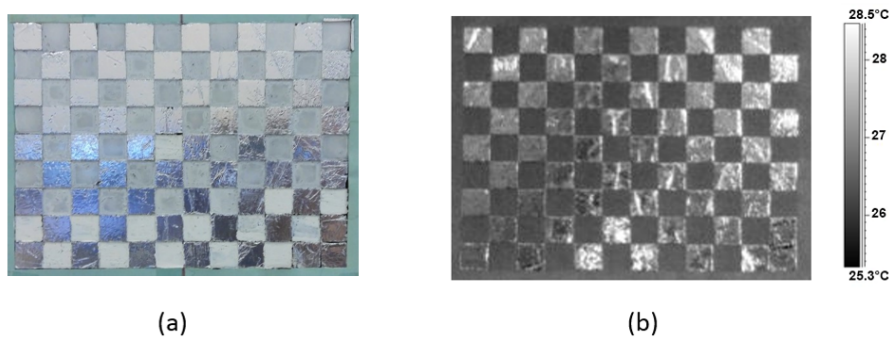
Fig. 1 – Visible (a) and thermal (b) image of the chessboard

A series of thermal images of the chessboard were acquired using a thermal camera FLIR SC655 (640×480 bolometer FPA, NETD <50 mK @ 30°C). More precisely, 5 thermal images of the chessboard were acquired, one for each of the 5 inclination angles (Fig. 2) at 3 different distances of acquisition (150 cm, 200 cm, 250 cm). The acquisition angles were -60°, -30°, 0°, 30° and 60° with respect to the vertical axis, corresponding to 0° (Fig. 2).
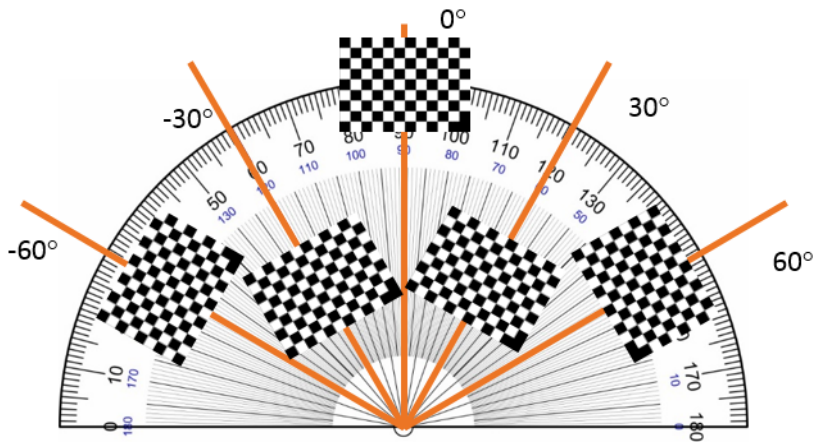
Fig. 2 – Inclination angles used for each of the 3 acquisition distances.

Therefore, the complete set of thermal images to warp was composed of 15 images (5 inclination angles (Fig. 2) x 3 acquisition distances).

The images were acquired in a completely thermal controlled room, with a stable temperature of 23°C.

As told, we did not control the acquisitions from thermal reflections and uniformity as we wanted to replicate, as much as possible, the real experimental conditions.

2.2 Participants

Fifteen users (age range: 30.13±7.36) took part in the study. All of them had no vision problem or correction through eyeglasses or contact lenses (11 out of 15). They were grouped in three categories according to their expertise with thermal IR imaging: i) high (more than 3 years of experience); ii) medium (more than 1 and less than 3 years of experience); iii) none (no experience) level of experience. They used a warping module software developed in Matlab (The MathWorks inc., Natick, MA, USA) and made available to the users in the form of an executable file. The warping module appeared as shown in Fig. 3. On the left side of the interface it was shown the raw thermal image to warp, whereas on the right side there was the reference checkerboard template with the numbered sequence of points.

For every loaded thermal image, the user had to select on the thermal image (Fig.3, left) the corresponding point shown on the template (Fig.3, right), for a total amount of 130 points per image and then press the CONFIRM pushbutton. The user had the possibility to zoom in and out, to pan the thermal image and to adjust the thermal contrast, through pushbuttons on the GUI. Thermal images were automatically loaded by the module and their order of presentation was randomized on distances and angles. The order of presentation was also randomized among users. The user could stop the procedure whenever he/she wanted and resume the procedure at a later time.
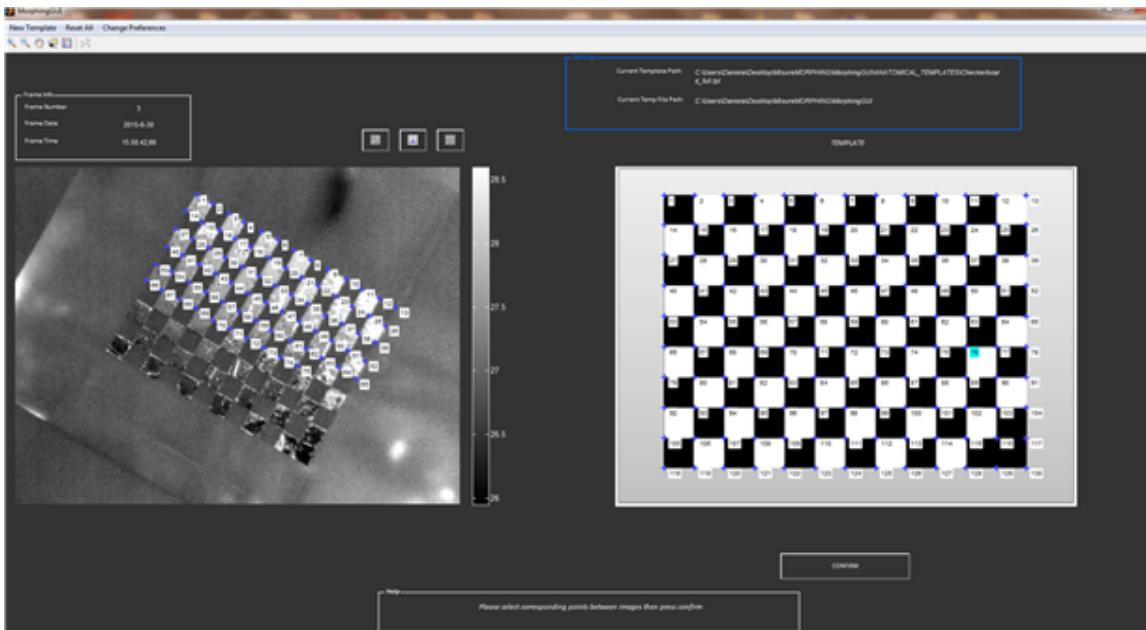
Fig. 3 – Warping module screenshot. On the left the thermal image to warp; on the right the template checkerboard with the numbered sequence of points

The subjects were asked to run the procedure at home. Therefore, the users' computers were not uniform in terms of screen resolution. In total, we had 7 users with 1920 x 1080 resolution and 8 with 1280 x 800.

2.3 Reference standard fiducial points.

To obtain a good and reliable estimation of the positioning error (relative to the manual positioning of points on the chessboard made by every user (OP points)) and of the thermal error (relative to the estimation in terms of temperature in each blob), it was necessary to establish a reference standard, to which every kind of measurement could be referred to. Specifically, we needed a reliable reference for the points' position in each of the fifteen thermal images used in the experimental protocol.

A semi-automated procedure was adopted to determine the reference standard positions of the fiducial points in each thermal image. A home-made script was developed using MatLab® 2013a. For each image, the binary map of the edges was first computed, using the Sobel edge detection algorithm with an automatic threshold level estimation. As it is a gradient-magnitude method, the default threshold value was chosen relatively to the zero-crossing method: edges were classified as large variations across zero [19].

Hough transform was then applied to the binary image, to identify all the single lines passing through the blobs' edges within the checkerboard (blue lines in Fig. 4 (a)) [20]. The intersection of these lines determined the fiducial points' position (red points in Fig. 4 (b)).

(a)   (b)
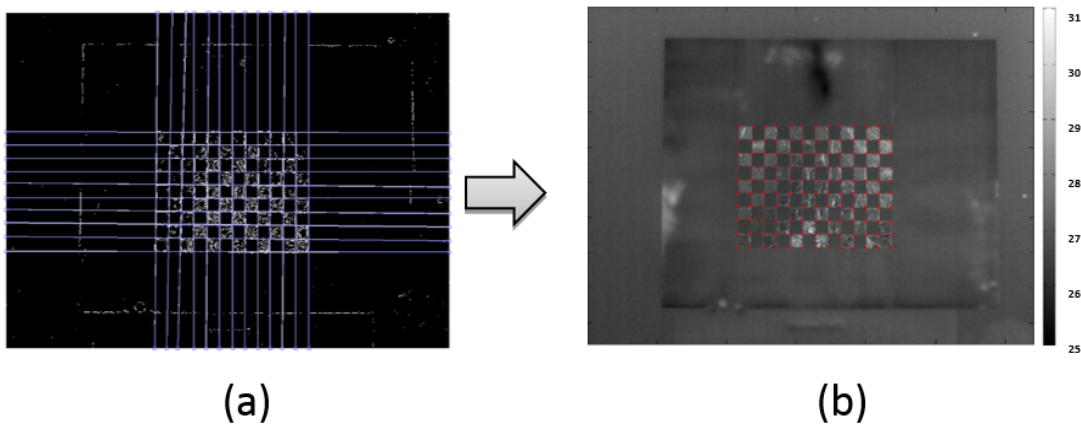
Fig. 4 – Determination of reference standard fiducial points. Detection of lines passing through the checkerboard, using Hough transform (a); reference standard fiducial points determined as intersection points of the lines (b)

The process is semi-automated because, if the automatic threshold evaluation failed in returning a good binary image in terms of edges detection, the threshold could be edited manually or through a slider (Fig. 5), in order to improve the lines detection through Hough transform.


(a)   (b)   (c)
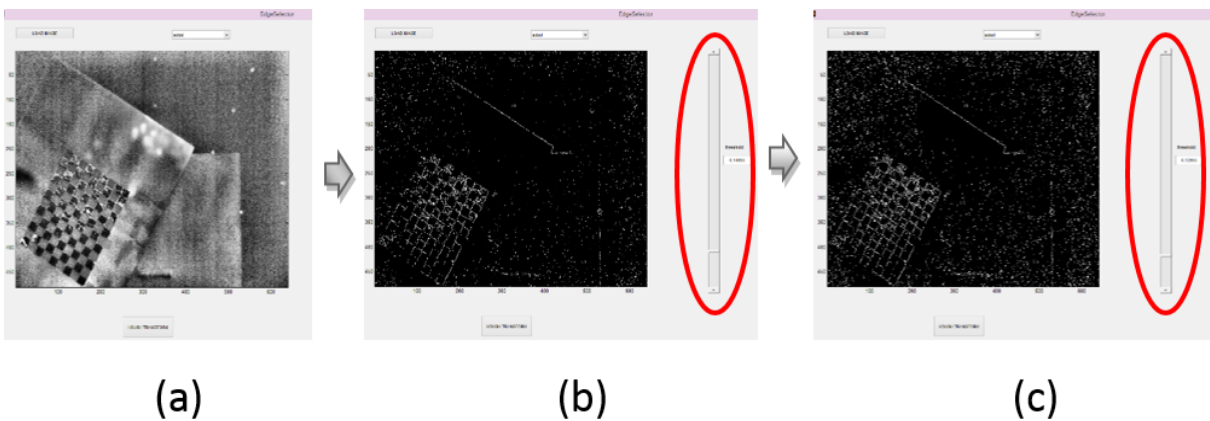
Fig. 5 – Semi-automated process of threshold determination using Sobel edge detection. Raw image (a); automatic threshold for Sobel edge detection (b); manually edited threshold through the slider (c)

Found false positive lines can be removed manually.

At the end of the whole procedure, for each of the 15 thermal images, we obtained 130 reference standard fiducial points.

The developed procedure is summarized in Fig. 6:
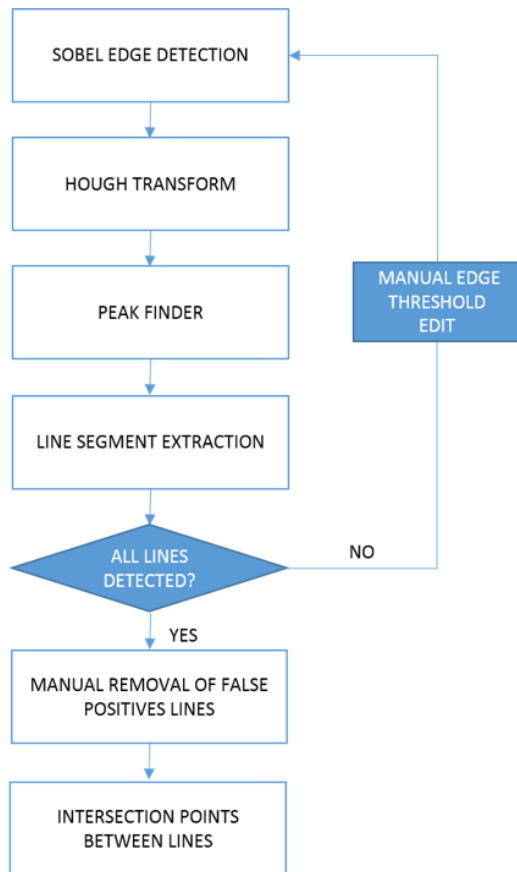
Fig. 6 – Flow chart for the reference standard feature's point determination

Repeated measures ANOVAs were performed to evaluate the accuracy of this method and to investigate the effect of the acquisition distances and inclination angles on the estimation of the position of the blob sides (both vertical and horizontal) through the reference standard.

2.4 Warping metrics: target registration error and thermal error evaluation

In order to evaluate the performance of the warping procedure on the thermal images, we estimated two categories of errors: the positioning error and the thermal errors. The former is known in literature as Target Registration Error (TRE) [21; 22; 23], while the latter was estimated through the elaboration of intensity errors described in [21; 23].

We estimated these two categories of errors with two purposes:

- evaluating the errors respect to the reference standard;
- evaluating the errors respect to the specific warping transformation used (i.e., warping mapping functions).

2.4.1 Errors between OP and REF patterns.

For every user, we determined the TRE for each of the acquired thermal images of the checkerboard (i.e. referring to the 5 inclination angles and 3 acquisition distances) for a total of 15 Images Under Test (IUTs) per user. This was done by calculating the Euclidean distance between the OP and the REF points for each IUT (Eq. 1).

$$\forall\ IUT, \forall\ point: \qquad TRE_{i,p}^* = \sqrt{(x_{OP}^{i,p} - x_{REF}^{i,p})^2 + (y_{OP}^{i,p} - y_{REF}^{i,p})^2}, \qquad\qquad\text{(Eq. 1)}$$

where $i \in [1,15]$ is the current IUT index, $p \in [1,130]$ is the current feature point, $(x_{OP}^{i,p}, y_{OP}^{i,p})$ is the coordinate of the *i-th* and *p-th* OP point and $(x_{REF}^{i,p}, y_{REF}^{i,p})$ is the coordinate of the *i-th* and *p-th* REF point.

For each of the 15 IUTs, we then averaged this value across the 130 points (Eq. 2):

$$\forall\ IUT: \qquad \overline{TRE}_i^* = \sum_{p=1}^{130} \frac{1}{130} TRE_{i,p}^* = \sum_{p=1}^{130} \frac{1}{130} \sqrt{(x_{OP}^{i,p} - x_{REF}^{i,p})^2 + (y_{OP}^{i,p} - y_{REF}^{i,p})^2}. \qquad\text{(Eq. 2)}$$

Concerning the thermal errors estimation, we evaluated the maximum, minimum, average and median temperature values for each blob of the checkerboard, for both OP and REF determined blobs. More precisely, for each blob of the chessboard, we extracted the maximum (T^max), minimum (T^min) and average (T^avg) values of the temperatures across all the pixels belonging to the specific blob and the median (T^mdn) value that is the central value of the temperature distribution of the pixels within that blob (i.e. the value separating the higher half of the thermal distribution from the lower half).

Each chessboard blob is identified by drawing a polyline connecting the 4 limit points (red line connecting Upper Left (UL), Upper Right (UR), Lower Left (LL) and Lower Right (LR) points), as shown in Fig. 7a. For the REF blobs, these points derived from the REF points, whereas for the OP blobs, the features points derived from OP points.
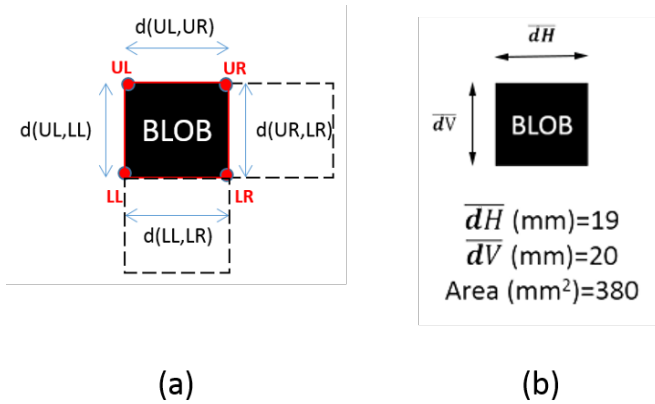

(a)   (b)

Fig. 7 – (a) Checkerboard blob definition: area inside the red line connecting Upper Left (UL), Upper Right (UR), Lower Left (LL) and Lower Right (LR) points; (b) Checkerboard blob dimensions: horizontal ($\overline{dH}$), vertical ($\overline{dV}$) and Area of the blob.

For each user and IUT, we evaluated the thermal errors for each blob calculating the difference between each of the above-mentioned parameters, for both OP and REF determined blob (Eq. 3-6):

$$\forall\ IUT, \forall\ blob: \qquad Thermal\ Error_{MAX}^* = T_{OP}^{max} - T_{REF}^{max}; \qquad\qquad\qquad\text{(Eq. 3)}$$

$$\forall\ IUT, \forall\ blob: \qquad Thermal\ Error_{MIN}^* = T_{OP}^{min} - T_{REF}^{min}; \qquad\qquad\qquad\text{(Eq. 4)}$$

$$\forall\ IUT, \forall\ blob: \qquad Thermal\ Error_{AVG}^* = T_{OP}^{avg} - T_{REF}^{avg}; \qquad\qquad\qquad\text{(Eq. 5)}$$

$$\forall\ IUT, \forall\ blob: \qquad Thermal\ Error_{MDN}^* = T_{OP}^{mdn} - T_{REF}^{mdn}. \qquad\qquad\qquad\text{(Eq.6)}$$

where $T_{OP}^{max}, T_{OP}^{min}, T_{OP}^{avg}, T_{OP}^{mdn}$ are the maximum, minimum, average and median thermal values for each blob respectively, determined through the OP points in each image and $T_{REF}^{max}, T_{REF}^{min}, T_{REF}^{avg}, T_{REF}^{mdn}$ are the maximum, minimum, average and median thermal values for each blob respectively, determined through the REF points in each image.

We, then, averaged these values across the 108 blobs of each IUT (Eq. 7-10):

$$\forall\, IUT: \quad \overline{Thermal\ error_{MAX}^*} = \frac{1}{108} Thermal\ Error_{MAX}^* = \frac{1}{108}(T_{OP}^{max} - T_{REF}^{max}); \qquad \text{(Eq. 7)}$$

$$\forall\, IUT: \quad \overline{Thermal\ error_{MIN}^*} = \frac{1}{108} Thermal\ Error_{MIN}^* = \frac{1}{108}(T_{OP}^{min} - T_{REF}^{min}); \qquad \text{(Eq. 8)}$$

$$\forall\, IUT: \quad \overline{Thermal\ error_{AVG}^*} = \frac{1}{108} Thermal\ Error_{AVG}^* = \frac{1}{108}\left(T_{OP}^{avg} - T_{REF}^{avg}\right); \qquad \text{(Eq. 9)}$$

$$\forall\, IUT: \quad \overline{Thermal\ error_{MDN}^*} = \frac{1}{108} Thermal\ Error_{MDN}^* = \frac{1}{108}\left(T_{OP}^{mdn} - T_{REF}^{mdn}\right). \qquad \text{(Eq. 10)}$$

Repeated measures ANOVAs were performed to assess the effect of the acquisition distances, inclination angles, and the users' level of experience on the positioning (TRE) and thermal errors.

2.4.2 Errors between raw IUT and transformed IUT.

We evaluated the positioning and thermal errors relative to both points and blobs of the raw IUT and points and blobs determined from the back-projected transformation of the template points in the IUT space (i.e. points represented as reference during warping procedure (Fig.3 – right image on the template chessboard)). Taking into account the BP template points, we aimed to evaluate the effect of the three warping transformations (affine, polynomial, LWM) on the errors. An example of the set of points involved in this particular procedure is shown in Fig. 8.
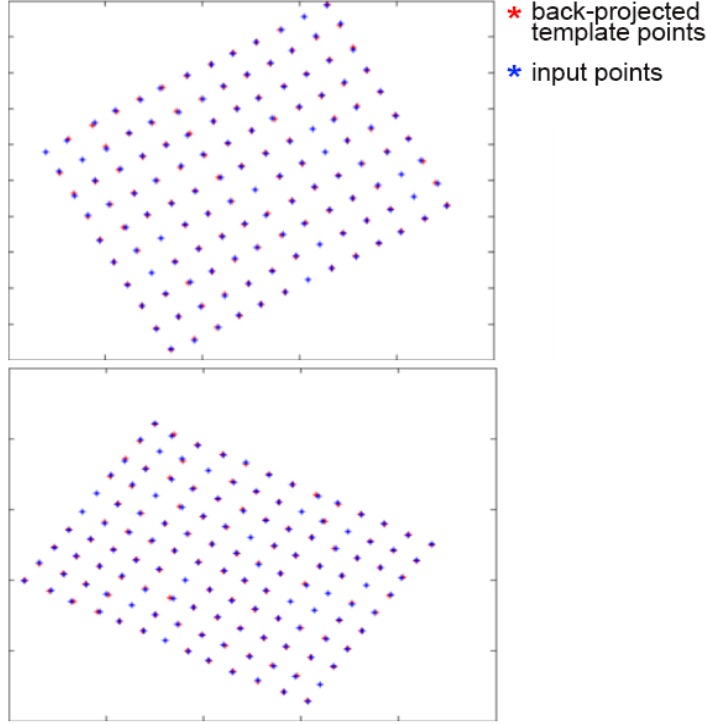
Fig. 8 – OP points (blue stars) and BP template points (red stars) for two representative images: @150cm, angle:-30° (upper image) and @200cm, angle: 30° (lower image)

For each of the three transformations, we determined the TRE for every user and IUT, calculating the Euclidean distance between OP points in the raw IUT and the BP template points on the IUT (Eq. 1 with $(x_{BP}^{i,p}, y_{BP}^{i,p})$ in place of $(x_{REF}^{i,p}, y_{REF}^{i,p})$). $(x_{BP}^{i,p}, y_{BP}^{i,p})$ is the coordinate of the *i-th* and *p-th* template point, back-projected on IUT space.

We, then, averaged this value for the 130 points (Eq. 2 with $(x_{BP}^{i,p}, y_{BP}^{i,p})$ in place of $(x_{REF}^{i,p}, y_{REF}^{i,p})$):

Concerning the thermal errors estimation, instead, we evaluated the maximum, minimum, average and median temperature values for each blob of the checkerboard, for both OP blobs and BP template blobs.

Each blob of the chessboard is determined by drawing a polyline connecting the 4 limit points (red line connecting Upper Left (UL), Upper Right (UR), Lower Left (LL) and Lower Right (LR) points), as shown above in Fig. 7.

For each user and IUT, we evaluated the thermal errors for each blob calculating the difference value for each of the above-mentioned parameter, between the OP- and BP-determined blobs (Eq. 3-6 with $T_{BP}^{max}, T_{BP}^{min}, T_{BP}^{avg}, T_{BP}^{mdn}$ in place of $T_{REF}^{max}, T_{REF}^{min}, T_{REF}^{avg}, T_{REF}^{mdn}$). $T_{BP}^{max}, T_{BP}^{min}, T_{BP}^{avg}, T_{BP}^{mdn}$ are the maximum, minimum, average and median thermal values for each blob, determined from the BP template points in each image respectively.

We, then, averaged these values across the 108 blobs within each IUT (Eq. 7-10 with $T_{BP}^{max}, T_{BP}^{min}, T_{BP}^{avg}, T_{BP}^{mdn}$ in place of $T_{REF}^{max}, T_{REF}^{min}, T_{REF}^{avg}, T_{REF}^{mdn}$).

Repeated measures ANOVAs were used to evaluate the effect of the warping transformations (i.e. affine, polynomial, LWM) on the positioning (TRE) and thermal errors.

3. Results

3.1 Reference standard accuracy

In order to check if the procedure to automatically determine the reference standard features points was reliable, we calculated the four sides of each blob in each image (horizontal measures: d(UL,UR), d(LL,LR) and vertical measures: d(UL,LL), d(UR,LR) in Fig. 7) referring to the reference standard features points pattern. We, then, averaged the computed values across the 108 blobs within the same image, obtaining the following results (Fig. 9):
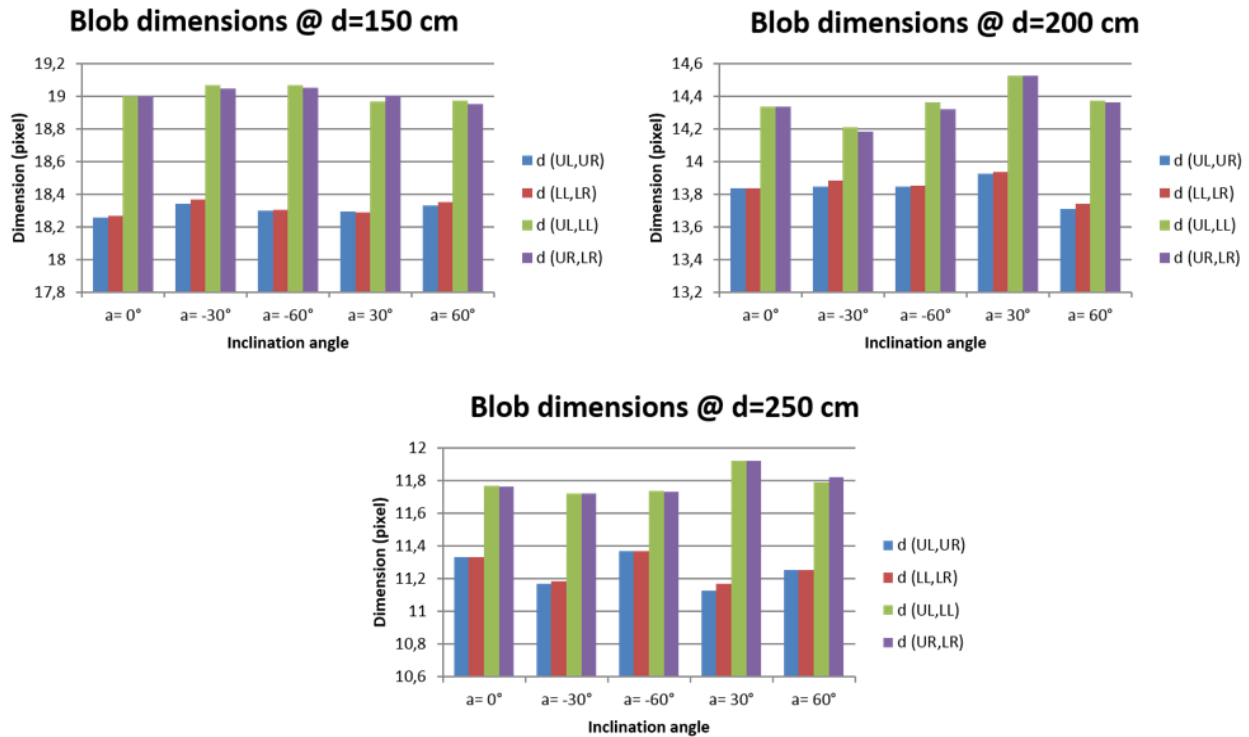

Fig. 9 – REF blob average dimensions (in pixel) for the different distances of acquisition: d=150 cm (a), d=200 cm (b), and d=250 cm (c). Horizontal dimensions are represented by blue and red bars, while vertical dimensions are shown in green and violet

First, it is possible to observe that the horizontal (red and blue bars in Fig. 9) and vertical dimensions (green and violet bars in Fig. 9) are slightly different. This result is in agreement with the real measurements on the checkerboard. We measured on the chessboard a mean horizontal blob dimension of $\overline{dH}$ =19 mm and a mean vertical blob dimension of $\overline{dV}$ =20 mm.

Moreover, in order to demonstrate the reliability of the REF, we performed repeated measures ANOVAs [24,25] for both the horizontal and vertical dimensions of the blobs.

Concerning the horizontal dimension, we obtained a significant difference *(i.e. statistical p<0.01)* only for the distance of acquisition *(F (2,106) = 7980.040, p<<0.01, η^2=0.993)*, as expected. Furthermore, the post-hoc test showed a significant difference in all the pairwise comparisons.

On the contrary, there were no significant differences related to the inclination angle *(F (4,104) = 0.151, p=0.962, η^2=0.006)* and to the position of the blob horizontal sides (i.e. upper vs. lower) *(F (1,107) = 3.477, p=0.065, η^2=0.031)*.

We achieved almost the same results for the vertical dimensions of the REF blobs. The repeated measures ANOVA showed a significant difference for the distance of acquisition (F $(2,106)$ = 12397.302, $p<<0.01$, $\eta^2=0.996$) and a slightly significant difference *(i.e. statistical $0.01<p<0.05$)* for the inclination angle (F $(4,104)$ = 3.558, $p=0.009$, $\eta^2=0.120$). However, the post-hoc tests revealed significant differences in all the pairwise comparisons for the acquisition distance and only one tendency to significance for the inclination angle when comparing -30° vs. 30° ($p=0.035$, $> p_{threshold}=0.01$).

Furthermore, there was no significant difference related to the position of the blob vertical sides (i.e. left vs. right) (F $(1,107)$ = 0.708, $p=0.402$, $\eta^2=0.007$).

The above-discussed repeated measures ANOVAs demonstrated that the horizontal and vertical dimensions of the blobs were invariant with inclination angles and that the horizontal upper and lower dimensions and vertical left and right dimensions were constant for each considered distance. Thus, based on these results, we could consider our reference standard as a reliable and accurate model.

Having established the REF model of our system, we were then able to perform a safe unit of measure conversion, from pixel to millimeters.

In fact, we calculated the average values of each of the blob dimensions (horizontal ($\overline{dH}$) and vertical ($\overline{dV}$), Fig. 7b) for each of the acquisition distances and each inclination angle, obtaining this conversion table (Table 1):

Acquisition distance	$\overline{dH}$ (pixel)	$\overline{dV}$ (pixel)	Area (pixel)	Conversion factor
150 cm	18.31	19.01	348.07	348.07/380 = **0.916 pixel/mm^2**
200 cm	13.84	14.35	198.60	198.60/380 = **0.522 pixel/ mm^2**
250 cm	11.26	11.79	132.75	132.75/380 = **0.349 pixel/ mm^2**

Table 1 - Conversion factor from pixel to millimetres at the different distances of acquisitions

3.2 Warping metrics

3.2.1 Statistics on positioning errors.

We performed statistical tests on the positioning error for the following contrasts: OP point positioning vs. REF features points and OP positioning vs. BP template points for the three warping transformations.

For both comparisons, we transformed the data units from pixel to millimeters, dividing the values in pixel by the conversion factors reported in the previous paragraph (Table 1) for each acquisition distance.

Concerning the comparison with REF features points, we performed repeated measures ANOVAs on the positioning errors. We obtained a significant difference for the distance of acquisition (*F (2, 13)* = 468.438, $p<<0.01$, $\eta^2=0.986$), as expected. Furthermore, the post-hoc test showed a significant difference in all the pairwise comparisons (Fig. 10).
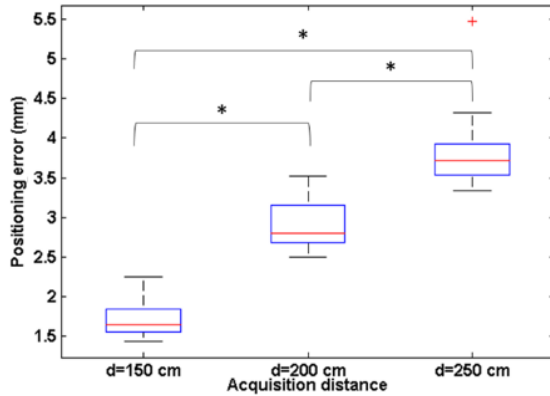
Fig. 10 – Box plot for positioning errors (OP vs. REF points) for the three acquisition distances

We obtained a slightly significant difference for the users' level of experience ($F (2, 13) = 5.315$, $p=0.021>p_{threshold}$, $\eta^2=0.450$). Furthermore, the post-hoc test showed a slightly significant difference in all the pairwise comparisons between high and medium ($p=0.035$) and medium and none ($p=0.015$) level of experience (Fig. 11).
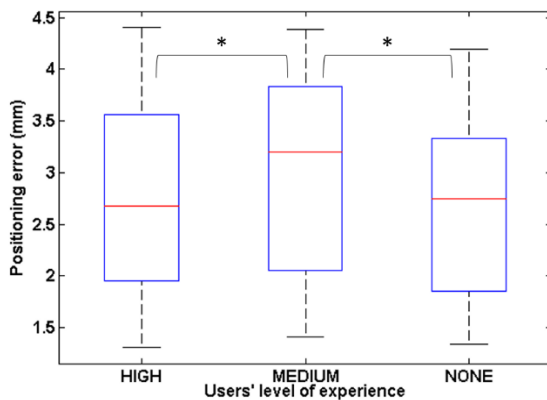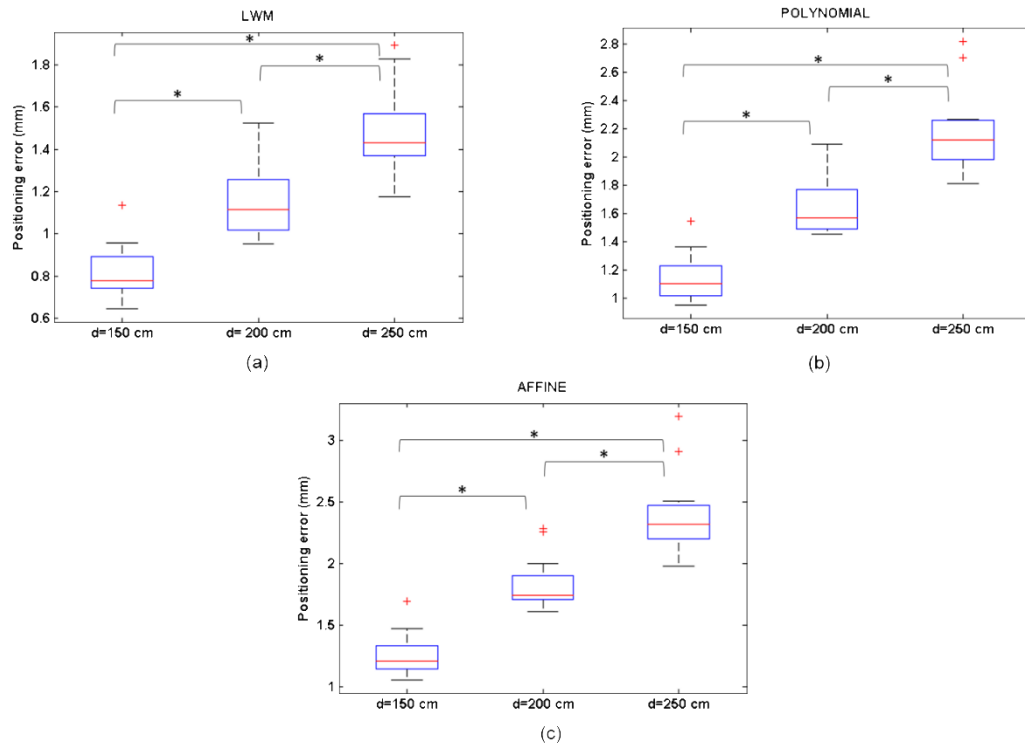

Fig. 11 – Box plot for positioning error (OP vs. REF points) for the three levels of user's experience

Repeated measures two-way ANOVA on the positioning error revealed a significant difference with respect to the distance x inclination angle ($F(8,7)=15.405$; $p<<0.01$; $\eta^2=0.946$) and post-hoc tests showed a significant difference for all the pairwise comparisons for the distances while for the inclination angle only for the following comparisons: 0° vs -30° ($p<<0.01$) and 0° vs 60° ($p<<0.01$) .

Concerning the comparison with the BP template point, we performed repeated measures ANOVA on the positioning errors. We found a significant difference for the distance of acquisition for the three transformations. In particular, for the LWM transformation we obtained $F(2,13)=327.858$; $p<<0.01$; $\eta^2=0.981$, for polynomial transformation we had $F(2,13)=325.684$; $p<<0.01$; $\eta^2=0.980$ and for affine transformation we obtained $F(2,13)=243.453$; $p<<0.01$; $\eta^2=0.974$.

In addition, the post-hoc tests showed a significant difference in all the pairwise comparisons for all the three transformation. Boxplots of the results are shown in Fig. 12.

Fig. 12 - Box plots for positioning error (OP vs. BP template points) for the three acquisition distances

We also compared the three transformation at a given acquisition distance through the repeated measures ANOVA obtaining significant differences. In particular, for d=150 cm: $F(2,13)=489.651$; $p<<0.01$; $\eta^2=0.987$; for d=200 cm: $F(2,13)=416.477$; $p<<0.01$; $\eta^2=0.985$; and for d=250 cm: $F(2,13)=201,497; p<<0.01; \eta^2=0.969$.

Furthermore, the post-hoc tests showed a significant difference in all the pairwise comparisons for all the three distances. Boxplots of the results are shown in Fig. 13.
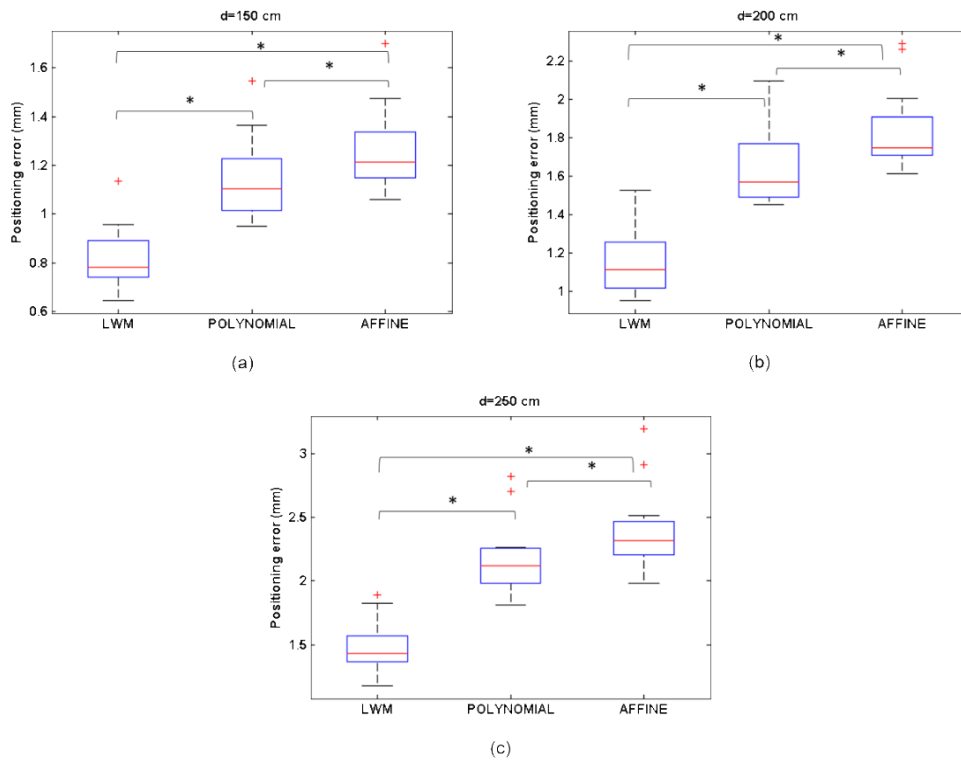
Fig. 13 - Box plots for positioning error (OP vs. BP template points) for the three warping transformations

3.2.2 Statistics on thermal errors.

We performed statistical tests on the thermal errors for the contrasts OP blobs vs. REF blobs parameters and OP blobs vs. BP template blobs parameters for the three warping transformations.

For the comparison with REF blobs, we performed repeated measures ANOVA on the thermal errors for average, median, maximum and minimum values of temperature within the blobs.

- ***Average temperature (T_{avg})***
  Repeated measures ANOVA on the thermal error on T_{avg} showed no significant difference with respect to the acquisition distance.
  Repeated measures ANOVA on the thermal error on T_{avg} revealed a significant difference with respect to the inclination angle ($F(4,11)=50.164$; $p<<0.01$; $\eta^2=0.948$) and post-hoc tests showed a significant difference for the following comparisons: slightly significant for -60° vs 30° ($p=0.011$) and 60° vs all inclination angles ($p<<0.01$) .
  No significant differences were found with respect to the experience level of the users.

- ***Median temperature (T_{mdn})***
  Repeated measures ANOVA on the thermal error on T_{mdn} showed a significant difference with respect to the distance of acquisition ($F(2,13)=45.076$; $p<<0.01$; $\eta^2=0.874$) and post-hoc tests revealed a statistically significant difference in the following pairwise comparisons: d=150 cm vs. d=200 cm ($p<<0.01$) and d=150 cm vs. d=250 cm ($p<<0.01$).
  Repeated measures ANOVA on the thermal error on T_{mdn} showed significant difference with respect to the inclination angle ($F(4,11)=140.091$; $p<<0.01$; $\eta^2=0.981$) and post-hoc tests showed a significant difference in the following pairwise comparisons: 0° vs. all inclination angles ($p<<0.01$ except 0° vs. -60°, where $p=0.012$) and 60° vs. all inclination angles ($p<<0.01$).

No significant differences were found with respect to the experience level of the users.

- ***Maximum temperature (T_{max})***

  Repeated measures ANOVA on the thermal error on T_{max} showed significant difference with respect to the distance of acquisition ($F(2,13)=29.604; p<<0.01; \eta^2=0.820$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the three distances.

  Repeated measures ANOVA on the thermal error on T_{max} showed a significant difference with respect to the inclination angle ($F(4,11)=181.859; p<<0.01; \eta^2=0.965$) and post-hoc tests showed a significant difference in the following pairwise comparisons: -60° vs. all inclination angles ($p<<0.01$).

  No significant differences were found with respect to the experience level of the users.

- ***Minimum temperature (T_{min})***

  Repeated measures ANOVA on the thermal error on T_{min} showed a significant difference with respect to the distance of acquisition ($F(2,13)=30.258; p<<0.01; \eta^2=0.823$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the three distances.

  Repeated measures ANOVA on the thermal error on T_{min} showed significant difference with respect to the inclination angle ($F(4,11)=89.482; p<<0.01; \eta^2=0.970$) and post-hoc tests showed a significant difference in the following pairwise comparisons: -60° vs. all inclination angles ($p<<0.01$), 0° vs.30° ($p=0.006$) and 60° vs. all inclination angles ($p<<0.01$).

  No significant differences were found with respect to the experience level of the users.
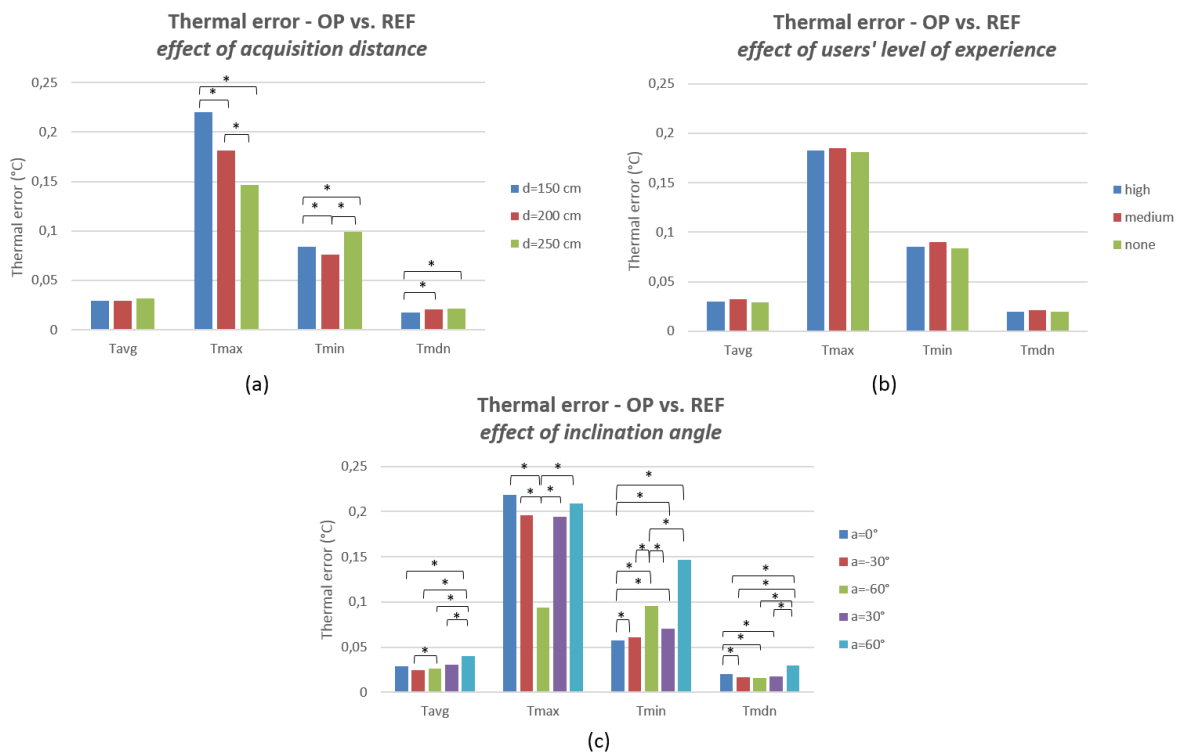
The above-mentioned results are summarized in Fig. 14:

Fig. 14 – Thermal errors between OP and REF blobs: effect of acquisition distance (a), users' level of experience (b) and inclination angle (c)

Concerning the comparison with BP template blobs, we performed repeated measurement ANOVAs on the thermal errors for average, median, maximum and minimum values of temperature in the blobs.

- **_Average temperature (T_{avg})_**
  Repeated measures ANOVA on the thermal error on T_{avg} showed a significant difference with respect to the acquisition distances for LWM transformation ($F(2,13)=19107.217$; $p<<0.01$; $\eta^2=1$), for polynomial ($F(2,13)=15862.940$; $p<<0.01$; $\eta^2=1$) as well as for affine transformation ($F(2,13)=15176.534$; $p<<0.01$; $\eta^2=1$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the acquisition distances.
  Repeated measures ANOVA on the thermal error on T_{avg} showed a significant difference with respect to the transformation ($F(2,13)=9.781$; $p<0.01$; $\eta^2=0.601$) and post-hoc tests revealed a significant difference in the pairwise comparison LWM vs. polynomial ($p=0.001$) and LWM vs. affine ($p=0.003$) transformation.
  Repeated measures ANOVA on the thermal error on T_{avg} showed a significant difference with respect to the inclination angle for LWM transformation ($F(4,11)=42794.410$; $p<<0.01$; $\eta^2=1$), for polynomial ($F(4,11)=26664.244$; $p<<0.01$; $\eta^2=1$) as well as for affine transformation ($F(4,11)=39507.376$; $p<<0.01$; $\eta^2=1$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the inclination angles.
  No significant differences were found respect to the experience level of the users.

- **_Median temperature (T_{mdn})_**
  Repeated measures ANOVA on the thermal error on T_{mdn} showed a significant difference with respect to the acquisition distances for LWM transformation ($F(2,13)=69952.248$; $p<<0.01$; $\eta^2=1$), for polynomial ($F(2,13)=56801.209$; $p<<0.01$; $\eta^2=1$) as well as for affine transformation ($F(2,13)=71676.701$; $p<<0.01$; $\eta^2=1$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the acquisition distances.
  Repeated measures ANOVA on the thermal error on T_{mdn} showed a significant difference with respect to the transformation ($F(2,13)=9.051$; $p<0.01$; $\eta^2=0.582$) and post-hoc tests revealed a significant difference in the pairwise comparison polynomial vs. affine ($p=0.003$) and slightly for LWM vs. affine ($p=0.037$) transformation.
  Repeated measures ANOVA on the thermal error on T_{mdn} showed a significant difference with respect to the inclination angle for LWM transformation ($F(4,11)=108730,216$; $p<<0.01$; $\eta^2=1$), for polynomial ($F(4,11)=58033.445$; $p<<0.01$; $\eta^2=1$) as well as for affine transformation ($F(4,11)=57300.173$; $p<<0.01$; $\eta^2=1$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the inclination angles.
  No significant differences were found with respect to the experience level of the users.

- **_Maximum temperature (T_{max})_**
  Repeated measures ANOVA on the thermal error on T_{max} showed a significant difference with respect to the acquisition distances for LWM transformation ($F(2,13)=144.584$; $p<<0.01$; $\eta^2=0.957$, for polynomial ($F(2,13)=134.495$; $p<<0.01$; $\eta^2=0.954$) as well as for affine transformation ($F(2,13)=154.895$; $p<<0.01$; $\eta^2=0.960$) and post-hoc tests revealed a significant difference in the pairwise comparisons d=150 cm vs. d=200 cm ($p<<0.01$) and d=200 cm vs. d=250 cm ($p<<0.01$).
  No significant differences were found with respect to the transformations.
  Repeated measures ANOVA on the thermal error on T_{max} showed a significant difference with respect to the inclination angle for LWM transformation ($F(4,11)=2996.437$; $p<<0.01$; $\eta^2=0.999$), for

polynomial ($F(4,11)=3361.778$; $p<<0.01$; $\eta^2=0.999$) as well as for affine transformation ($F(4,11)=1972.857$; $p<<0.01$; $\eta^2=0.999$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the inclination angles.

Repeated measures ANOVA on the thermal error on T_{max} showed a significant difference with respect to the users' level of experience only for LWM transformation ($F(2,13)=23,480$; $p<<0.01$; $\eta^2=0.783$). In this case, post-hoc tests showed a significant difference in the pairwise comparison high vs. none ($p<<0.01$) and medium vs. none ($p=0.004$) level of experience.

- **Minimum temperature (T_{min})**

    Repeated measures ANOVA on the thermal error on T_{min} showed a significant difference with respect to the acquisition distances for LWM transformation ($F(2,13)=696.921$; $p<<0.01$; $\eta^2=0.991$, for polynomial ($F(2,13)=676.629$; $p<<0.01$; $\eta^2=0.990$) as well as for affine transformation ($F(2,13)=842.978$; $p<<0.01$; $\eta^2=0.992$) and post-hoc tests revealed a significant difference in all the pairwise comparisons for all the acquisition distances.

    Slightly significant difference was reported from repeated measures ANOVA with respect to the transformation ($F(2,13)=4.748$; $p=0.028$; $\eta^2=0.422$) and post-hoc tests showed a tendency to a significant difference in the pairwise comparison polynomial vs. affine ($p=0.059$) transformation.

    Repeated measures ANOVA on the thermal error on T_{min} showed significant difference with respect to the inclination angle for LWM transformation ($F(4,11)=1414.521$; $p<<0.01$; $\eta^2=0.998$), for polynomial ($F(4,11)=1631.267$; $p<<0.01$; $\eta^2=0.998$) as well as for affine transformation ($F(4,11)=3955.224$; $p<<0.01$; $\eta^2=0.999$) and post-hoc tests showed a significant difference in all the pairwise comparisons for all the inclination angles (with a slightly significant difference for comparison a=-30° vs a=-60° ($p=0.022$) for LWM) except for comparison a=-30° vs a=-60° for both polynomial and affine transformations.

The above-mentioned results are summarized in Fig. 15 for LWM, Fig. 16 for polynomial, and Fig. 17 for affine transformations:
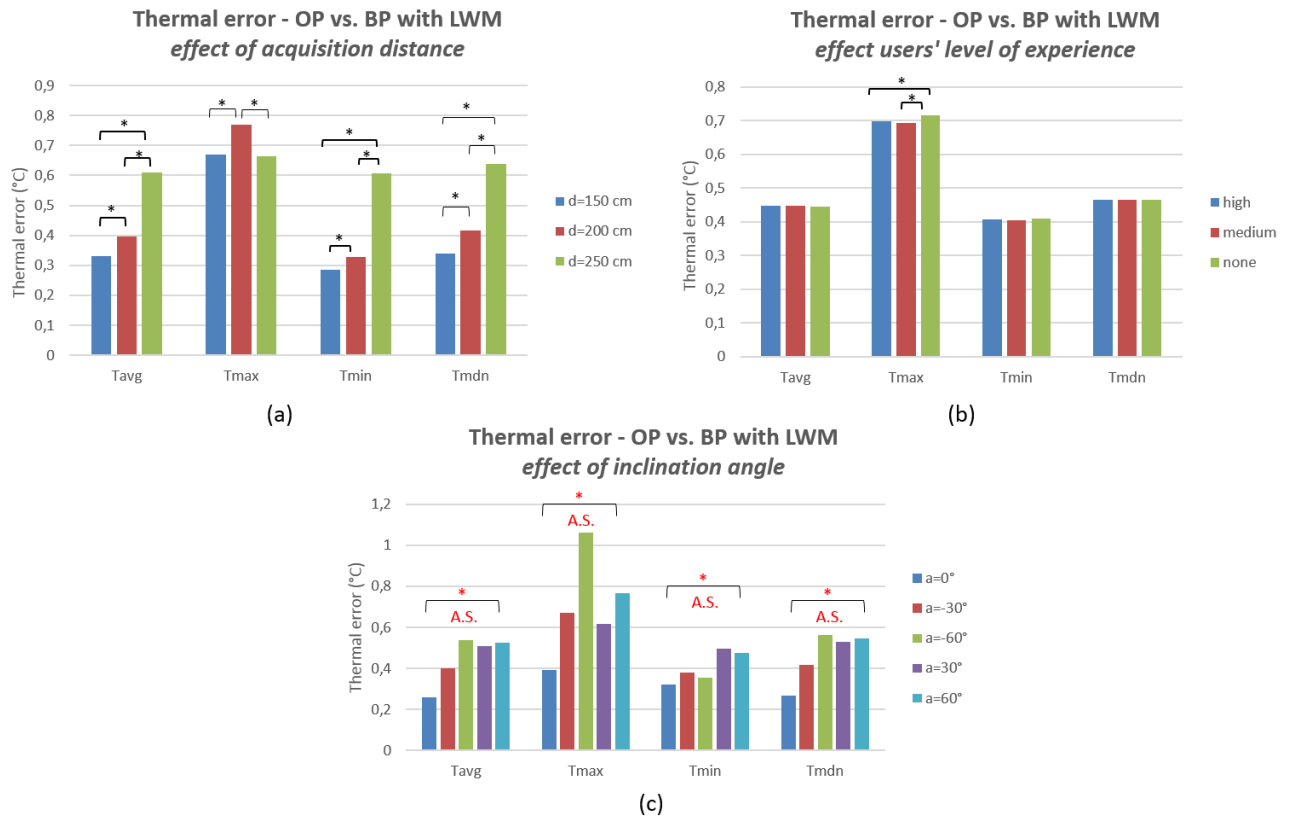
Fig. 15 - Thermal errors between OP and BP blobs: effect of acquisition distance (a), users' level of experience (b) and inclination angle (c) for LWM transformation (A.S.= all significant).

Fig. 16 - Thermal errors between OP and BP blobs: effect of acquisition distance (a), users' level of experience (b) and inclination angle (c) for POLYNOMIAL transformation (A.S.= all significant, N.S.= non-significant).
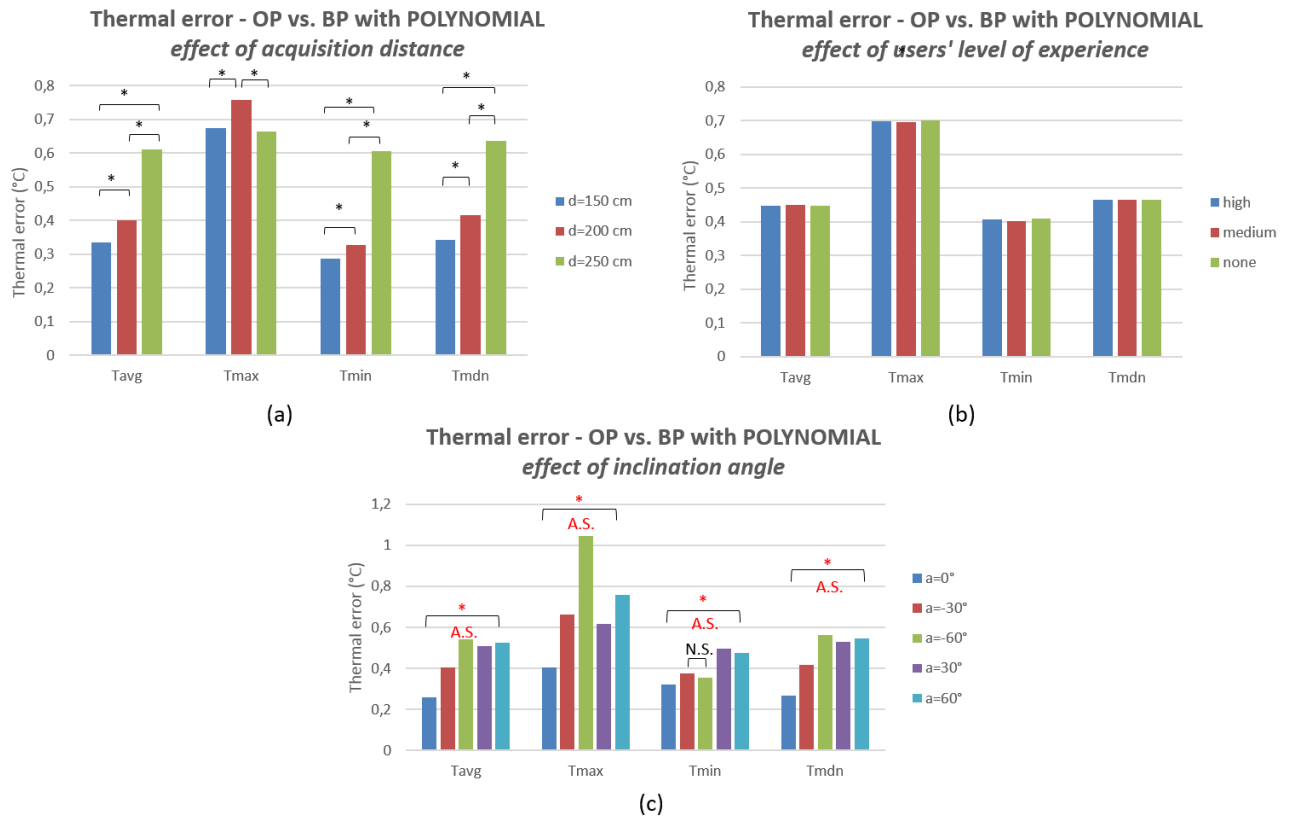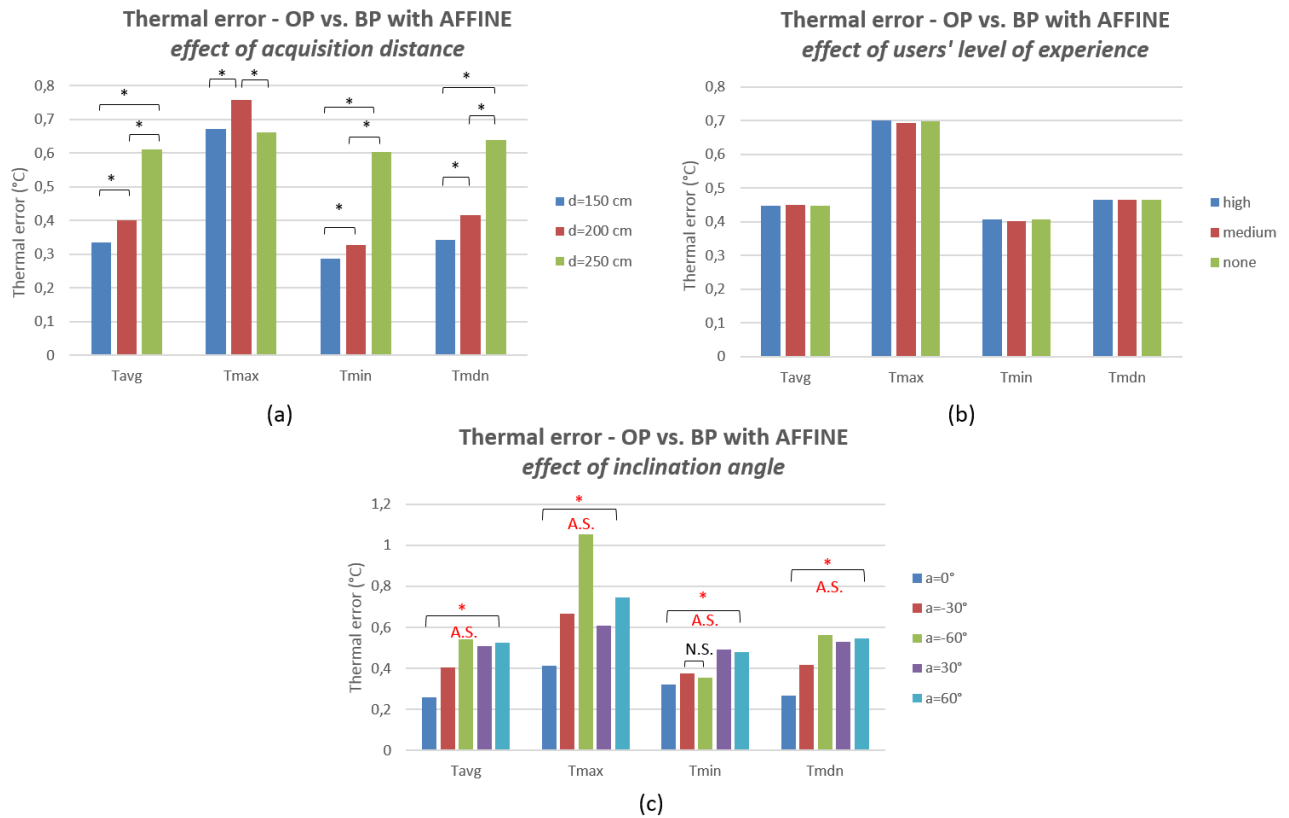
Fig. 17 - Thermal errors between OP and BP blobs: effect of acquisition distance (a), users' level of experience (b) and inclination angle (c) for AFFINE transformation (A.S.= all significant, N.S.= non-significant).

Fig. 18 summarizes the results for the comparisons among the transformations for the four investigated thermal errors:
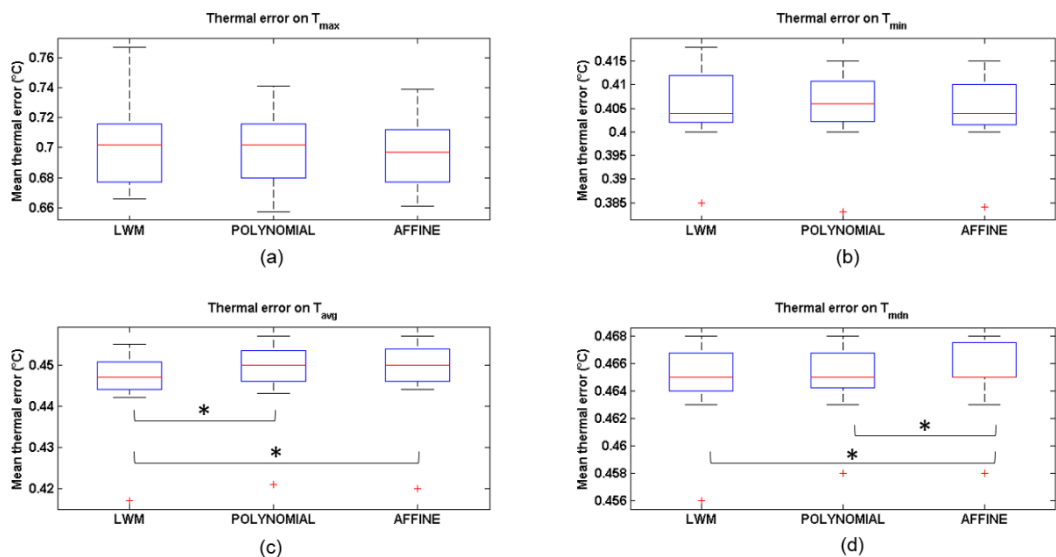

Fig. 18 - Thermal errors between OP and BP blobs: comparison among transformation of the thermal errors for maximum (a), minimum (b), average (c) and median (d) temperature

Considerations on different screen resolutions among users.

The users used two different screen resolutions: 7 users with a 1920 x 1080 and 8 with a 1280 x 800 resolutions.

In order to check if the screen resolution had an influence in our experiment results, we performed a Mann–Whitney U test on the positioning errors between the two samples.

None of the two comparison (OP vs. REF points and OP vs. BP template points) were significant.

Discussion and Conclusions

This work aimed at demonstrating the feasibility of applying a warping procedure on thermal IR images.

To this goal, we acquired thermal images of a home-made thermal checkerboard at three different acquisition distances (d=150 cm, d=200 cm, d=250 cm) and at five different inclination angles (0°, -30°, -60°, 30°, 60° with respect to the vertical axis (Fig.2)), for a total of 15 IUTs. Users with different level of experience (high, medium, and none) in thermal IR imaging technique had to manually select a set of 130 points on each thermal chessboard image, following the positioning order indicated on a reference template (Fig.3).

After having established a reference standard on the checkerboard images (see Section 2.3), we evaluated the effect of the inclination angles, acquisition distances, and users' level of experience on the positioning and thermal errors computed between the OP and REF set of points. In addition, we investigated the same effects on the positioning and thermal errors between the OP and BP template points calculated with three different warping transformations (LWM, polynomial and affine).

Below, we summarize and discuss the results from four different points of view:

1. Concerning the **_acquisition distance_**, we found that -as expected- it had the strongest effects on both positioning and thermal errors. Indeed, we noticed that the positioning error between OP and REF set of points increased with the acquisition distance (Fig. 11) as well as the positioning error between OP and BP template points for all the three transformations (Fig. 13). In particular, we observed that the lowest positioning error is made at lower distance (i.e., d=150 cm). The acquisition distance has a strong influence also on thermal errors. In fact, positioning errors on T_{avg}, T_{min}, and T_{mdn} errors -but not T_{max}- increased with the acquisition distance for all the three transformations (Fig. 16 (a), Fig. 17 (a), and Fig. 18 (a)). Also for the contrast OP vs. REF blobs (Fig. 15 (a)), thermal errors for T_{avg} and T_{mdn} –but not T_{max} and T_{min}- increased with the acquisition distance.
   The strong effect of the acquisition distance is related to a better resolution of the thermal images for shorter distances respect to the camera. In fact, the image resolution increases as the camera is closer to the object, so that borders and details can be more easily detected by the user respect to far images.
   Moreover, the effect of the acquisition distance is related also to the nature of the used technique. In fact, IR thermal imaging quantitatively measures radiative part of total heat flux, spontaneously emitted by the body: the closer is the body to the thermal camera, the higher amount of signal is detected.

2. Concerning the **level of experience**, we noticed that it does not strongly affect either the positioning and thermal errors. In fact, we only found a slightly significant difference on the positioning error for the comparison OP vs REF points for high vs medium level of experience and medium vs none experience (Fig. 12). In terms of thermal errors, we only found a significant difference of the level of experience on T_{max} for the contrast OP vs BP blobs for the LWM transformation. In this case, users with higher experience were statistically different from the other two groups (Fig. 16 (b)).

   The small effect of the users' level of experience on the warping performance demonstrates that the warping procedure is an objective method that can be easily and effectively used also from naïve users. However, this might by related to the geometrical regularity and structural simplicity of the chessboard, where adjacent blobs were clearly recognizable and points were not hard to detect.

3. Concerning the **inclination angle**, we could not find specific and reproducible effects on the errors (probably because we did not control the acquisitions from thermal reflections and uniformity, replicating, as much as possible, the real experimental conditions) except for the comparison between OP and BP blobs respect to the three transformations. Here we observed the lowest thermal error on the 0° thermal image for T_{avg}, T_{min}, T_{max} and T_{mdn} (Fig. 16 (c), Fig. 17 (c), and Fig. 18 (c)).

   The small effect of the inclination angle further demonstrated that the warping procedure in an easy-to-apply method, even with different inclinations of the observed target. However, results suggest that to minimize thermal errors, 0° target inclinations are preferred.

4. Concerning the comparison between the three kind of **warping transformations**, we observed that the lowest positioning error is reached with the LWM transformation for the all the three acquisition distances (Fig. 14). We found an effect of the warping transformation also on the thermal errors between OP and BP blobs, in particular for T_{avg} and T_{mdn} errors (Fig. 19 (c-d)) but not for T_{max} and T_{min} (Fig. 19 (a-b)).

   This is probably due to the different nature of LWM with respect to the polynomial or affine transformations. The former is a local transformation that locally acts around the points, while the last two are global transformations operating on the whole image, without giving a different weight to the different parts of the image.

5. Concerning the **screen resolution**, we found that it does not influence the performances of the thermal warping procedure if we compare higher screen resolutions (1920 x 1080) with lower ones (1280 x 800).

   This is probably due to the possibility of users to zoom in and out the images while processing them, counterbalancing the effect of a poorer screen resolution with an enlargement of the image size and vice versa.

To the best of our knowledge, this work represents the first attempt to investigate from both a technical and a quantitative point of view the feasibility of the warping approach on IR thermal imaging. Even though the results reported in this study are preliminary and further improvements are needed, they could represent the starting point for future and more complicated warping applications. In future works, we plan to recruit a larger sample of users and to study more complex patterns to warp with the aim of developing an accurate procedure for warping human faces or bodies on anatomical templates.

References

1. Aldave, I. J., Bosom, P. V., González, L. V., De Santiago, I. L., Vollheim, B., Krausz, L., & Georges, M. (2013). Review of thermal imaging systems in composite defect detection. *Infrared Physics & Technology*, *61*, 167-175.
2. Talairach J, Tournoux P (1988). Co-planar stereotaxic atlas of the human brain. Thieme, New York.
3. Ring, E., Ammer, K., Jung, A., Murawski, P., Wiecek, B., Zuber, J., Zwolenik, S., Plassmann, P., Jones, C., & Jones, B. F. (2003, December). Standardization of infrared imaging. In *Conference proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* (Vol. 2, pp. 1183-1185).
4. Toga, A. W. (Ed.). (1998). *Brain warping*. Academic press.
5. Glasbey, C. A., & Mardia, K. V. (1998). A review of image-warping methods. *Journal of applied statistics*, *25*(2), 155-171.
6. Brown, L. G. (1992). A survey of image registration techniques. *ACM computing surveys (CSUR)*, *24*(4), 325-376.
7. Tang, Y. Y., & Suen, C. Y. (1993). Image transformation approach to nonlinear shape restoration. *Systems, Man and Cybernetics, IEEE Transactions on*, *23*(1), 155-172.
8. Heikkila, J., & Silvén, O. (1997, June). A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on* (pp. 1106-1112). IEEE.
9. Maurer, C. R., & Fitzpatrick, J. M. (1993). A review of medical image registration. *Interactive image-guided neurosurgery*, *17*.
10. Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, *21*(11), 977-1000.
11. Schaefer, G., Tait, R., & Zhu, S. Y. (2006, August). Overlay of thermal and visual medical images using skin detection and image registration. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE* (pp. 965-967). IEEE.
12. Toet, A., Van Ruyven, L. J., & Valeton, J. M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering*, *28*(7), 287789-287789.
13. Istenic, R., Heric, D., Ribaric, S., & Zazula, D. (2007, June). Thermal and visual image registration in hough parameter space. In *Systems, Signals and Image Processing, 2007 and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services. 14th International Workshop on* (pp. 106-109). IEEE.
14. Howell, K. J., Lavorato, A., Visentin, M. T., Smith, R. E., Schaefer, G., Jones, C. D., Weibel, L., Denton, C.P., Harper, J., & Woo, P. (2009). Validation of a protocol for the assessment of skin temperature and blood flow in childhood localised scleroderma. *Skin Research and Technology*, *15*(3), 346-356.
15. Kong, S. G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B. R., Koschan, A., Mingzhong, Y., & Abidi, M. A. (2007). Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. *International Journal of Computer Vision*, *71*(2), 215-233.
16. Mohd, M. N. H., Kashima, M., Sato, K., & Watanabe, M. (2014). Facial visual-infrared stereo vision fusion measurement as an alternative for physiological measurement. *Journal of Biomedical Image Processing (JBIP)*, *1*(1), 34-44.
17. Cheng, T. Y., & Herman, C. (2013, March). Involuntary motion tracking for medical dynamic infrared thermography using a template-based algorithm. In *SPIE Medical Imaging* (pp. 86692Q-86692Q). International Society for Optics and Photonics.
18. Vollmer, M., & Möllmann, K. P. (2010). *Infrared thermal imaging: fundamentals, research and applications*. John Wiley & Sons.
19. Sobel, I. (1978). Neighborhood coding of binary images for fast contour following and general binary array processing. *Computer graphics and image processing*, *8*(1), 127-135
20. Hough, P. V. (1962). *Method and means for recognizing complex patterns* (No. US 3069654).
21. Fitzpatrick, J. M., Hill, D. L., Shyr, Y., West, J., Studholme, C., & Maurer, C. R. (1998). Visual assessment of the accuracy of retrospective registration of MR and CT images of the brain. *Medical Imaging, IEEE Transactions on*, *17*(4), 571-585.

22. Latifi, K., Zhang, G., Stawicki, M., van Elmpt, W., Dekker, A., & Forster, K. (2013). Validation of three deformable image registration algorithms for the thorax. *Journal of Applied Clinical Medical Physics*, *14*(1).

23. Yin, L., Tang, L., Hamarneh, G., Celler, A., Shcherbinin, S., Fua, T., Carolan, H., Thompson, A., Liu, M., Duzenli, C., Gill, B., Sheehan, F., Powe, J., Worsley, D., & Moiseenko, V. (2009). Evaluating geometrical accuracy of image registration methods in SPECT guided radiation therapy. In *Medical Image Computing and Computer-Assisted Intervention Workshop on Geometric accuracy in image guided intervention (MICCAI GAIGI)* (pp. 63-70).

24. Von Ende, C. N. (2001). Repeated-measures analysis. *Design and analysis of ecological experiments. Oxford University Press, Oxford*, 134-157.

25. Girden, E. R. (1992). *ANOVA: Repeated measures* (No. 84). Sage.