

Article

Distance-Dependent Multimodal Image Registration for Agriculture Tasks

Ron Berenstein ^{1,*}, Marko Hočevar ^{2,†}, Tone Godeša ^{2,†}, Yael Edan ^{1,†} and Ohad Ben-Shahar ^{3,†}

¹ Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel; E-Mail: yael@bgu.ac.il

² Faculty of Mechanical Engineering, University of Ljubljana, Aškerčeva 6, Ljubljana 1000, Slovenia; E-Mails: marko.hocevar@fs.uni-lj.si (M.H.); tone.godesa@kis.si (T.G.)

³ Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel; E-Mail: ben-shahar@cs.bgu.ac.il

† These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: berensti@bgu.ac.il; Tel.: +972-50-6220657.

Academic Editor: Vittorio M. N. Passaro

Received: 1 May 2015 / Accepted: 11 August 2015 / Published: 21 August 2015

Abstract: Image registration is the process of aligning two or more images of the same scene taken at different times; from different viewpoints; and/or by different sensors. This research focuses on developing a practical method for automatic image registration for agricultural systems that use multimodal sensory systems and operate in natural environments. While not limited to any particular modalities; here we focus on systems with visual and thermal sensory inputs. Our approach is based on pre-calibrating a distance-dependent transformation matrix (DDTM) between the sensors; and representing it in a compact way by regressing the distance-dependent coefficients as distance-dependent functions. The DDTM is measured by calculating a projective transformation matrix for varying distances between the sensors and possible targets. To do so we designed a unique experimental setup including unique Artificial Control Points (ACPs) and their detection algorithms for the two sensors. We demonstrate the utility of our approach using different experiments and evaluation criteria.

Keywords: sensor registration; control points; artificial control points; sensor fusion

1. Introduction

Implementing accurate, selective and economical agricultural robots for various agricultural and horticultural tasks is one of the main goals of precision agriculture (see Table 1). A critical task within this field is the detection of natural objects, a particularly difficult challenge due to the inherent unconstrained environments and the significant variability in all object properties such as color, shape, size, texture, and reflectance properties. Moreover, the highly unstructured scenes are often characterized by large degree of uncertainty, changing illumination and shadow conditions, and severe occlusions. Combined with the sheer complexity of the typical unstructured agricultural scene, it is clear why this domain is considered one of the ultimate challenges of sensory systems, machine vision systems in particular [1].

State-of-the-art fruit detection systems commonly combine several detection sensors and algorithms using sensor fusion techniques to achieve better detection rates (Table 1). Often, the detection module constitutes a single RGB camera combined with one or more complementary sensors such as thermal, infra-red, laser scanner, hyper-spectral, or time of flight (see Table 1 for many examples). To do so successfully, the first mandatory step in fusing the data from the different sensors is the registration of their images. This registration is the focus of our work. While nothing in our approach is tailored to any particular sensor modalities, our case study is the combination of visual and thermal sensors. Indeed, how fruit detection rates can be improved by fusing these particular sensors compared to each of them alone was already demonstrated in the past, for example for the detection of oranges [2].

Thermal imaging allows for finding fruit whose color, potentially an indicative cue, is very sensitive to illumination conditions. Leaves, unlike fruits, accumulate less heat and emit it for a shorter time. However, since the thermal response is sensitive to the sunlight illumination and heat accumulation, fruit on different parts of the tree might respond differently [1–3]. In such cases combining thermal and RGB imaging could be useful for fruit detection and harvesting, under the assumption that a suitable registration procedure is provided.

Table 1. Examples of various precision agriculture topics of research in the last three decades.

Task	Sensors	Description of Research	Reference
Target detection	RGB, spectral	survey on fruit detection using computer vision	[4]
	RGB camera	detection of grape clusters	[5]
	Thermal vision	detection of oranges using thermal imagery	[6]
	RGB camera + thermal camera	detection of oranges using RGB camera and thermal camera and conducting fusion between	[2]

Table 1. Cont.

Task	Sensors	Description of Research	Reference
Precision spraying	RGB camera	selective sprayer for weed control using machine vision, real time controller and controllable spraying system	[7]
	CCD camera	developed a tree crown recognition and smart spraying system	[8]
	BW camera + ultrasonic	speed spraying using fuzzy logic control of machine vision and ultrasonic sensors	[9]
	Ultrasonic	spraying robot for vine production	[10]
Robotic harvesting	Stereo vision	apple harvester robot	[11]
	Color camera (HSV)	robot for greenhouse operations	[12]
	Color camera	apple harvester robot	[13]
	2 wavelengths of color camera	strawberry harvester robot	[14]

Image registration is the process of overlaying (*i.e.*, transforming into the same coordinate system) two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors [15]. Registration processes are usually divided into the following steps: (i) **Feature detection**—detects (manually or automatically) the position of distinctive objects in the image, also known as Control Points (CPs). While such features are often the starting point for the registration process [16–19], in many cases they are replaced with pixel patches that provide distinct appearance [20–22]; (ii) **Feature matching**—establishes correspondence between the CPs of the different sensors. For example, in Figure 1 we may wish to correspond the yellow disk in the visual image with the left most point in the thermal image; (iii) **Transformation model estimation**—determines the type of image transformation according to prior information (or lack thereof) regarding the acquisition process; (iv) **Image transformation**—transforms the two images into a common coordinate system (often by transforming one into the other).

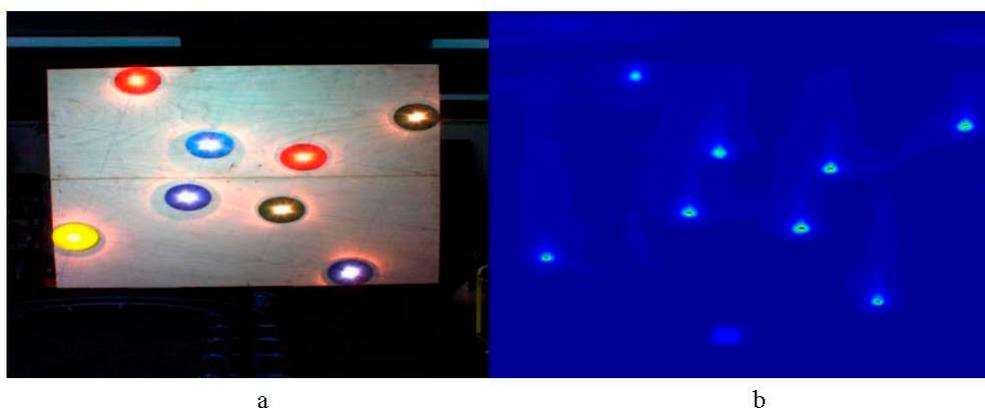


Figure 1. Same scene acquired by different sensors. (a) Image from RGB camera; (b) Image from thermal camera.

Applying automatic image registration for systems that use visual and thermal images is difficult due to the different methods employed to calculate the CPs for each type of sensor and the lack of natural

correspondence between their CPs [23]. Most papers dealing with visual and thermal image fusion and registration [23,24] perform this correspondence manually (*i.e.*, use a human to mark corresponding CPs in the visual and thermal images). One exception is Istenic *et al.* [25] who presented an automatic method for visual and thermal image registration. Their approach uses edge detection and Hough transform to detect linear lines in both images and to compute CPs for the registration process. However, since this approach is based on straight linear segments, it is limited only to images that consist of objects with straight lines [25], which are not prevailing in the unstructured agriculture environment, another method must be developed. Introducing landmarks manually for thermal and visible image registration proved successful for aligning static medical images [26,27] and for fruit registration (e.g., [2]).

Among the mentioned papers [23,24,26,27] require manual CP selection. The procedure for Istenic *et al.* [25] is automatic, however their scenes were much more structured.

The goal of this work is to introduce a new approach for highly accurate registration of thermal and color images.

The novelty of the manuscript is that it offers a new registration method suitable for unstructured environments with long intervals of sensing ranges. It introduces a new approach for highly accurate registration of thermal and color images. The registration approach is based on the computation of a “dynamic” transformation matrix (TM) in which each element is a function of the distance from the object in the image. In the field this distance can be measured by a range sensor. In this paper, we demonstrate the utility of our approach on a robotic sprayer equipped with an RGB and thermal camera, as well as a laser scanner. Within the interval, the method offers compact representation of multiple (or infinite, if one considers the continuous range) registration transformations. Thanks to the regression algorithm, the procedure permits registration at distances for which the sensors were not calibrated for. For completeness we also present the ACP design and algorithm for ACP detection, although this is not the core of the paper.

2. Methods

2.1. Image Registration Model

The core of the registration process is to find TM that transforms pixels in one image to the other. Formally,

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = TM \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}; \quad TM = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \quad (1)$$

where x' and y' is the registered pixel coordinates and x, y is the original pixel coordinates. To be general enough, here we assume that the transformation between the multimodal images is projective, an eight DOF (Degree Of Freedom) mapping which contributes to the registration accuracy [28].

The projective transformation is a 3×3 homogeneous matrix (TM). Since it has eight DOFs, a linear system of rank eight is required in order to determine the value for each of the TM elements. Each CP (identified by its two coordinates) can contribute two constraints by applying Equation (1) and rearranging as follows

$$\begin{aligned}
 x' &= \frac{h_{00}x + h_{01}y + h_{02}}{h_{20}x + h_{21}y + h_{22}} & y' &= \frac{h_{10}x + h_{11}y + h_{12}}{h_{20}x + h_{21}y + h_{22}} \\
 h_{00}x + h_{01}y + h_{02} - h_{20}xx' - h_{21}yx' - h_{22}x' &= 0 \\
 h_{10}x + h_{11}y + h_{12} - h_{20}xy' - h_{21}yy' - h_{22}y' &= 0
 \end{aligned}
 \tag{2}$$

By repeating this procedure with at least four CP, a full linear system can be created and solved in order to determine the values of the TM.

$$\begin{bmatrix}
 x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 \\
 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'x_1 & -y_1'y_1 \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 & & & & & & & \\
 x_n & y_n & 1 & 0 & 0 & 0 & -x_n'x_n & -x_n'y_n \\
 0 & 0 & 0 & x_n & y_n & 1 & -y_n'x_n & -y_n'y_n
 \end{bmatrix} \cdot \begin{bmatrix} h_{00} \\ h_{01} \\ h_{02} \\ h_{10} \\ h_{11} \\ h_{12} \\ h_{20} \\ h_{21} \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ \vdots \\ x_n' \\ y_n' \end{bmatrix}
 \tag{3}$$

Four CPs is the minimum required to solve the projective transformation matrix (TM) but using more CPs can provide resistance to noise and contribute to the registration accuracy. In this case the system in Equation (3) becomes overdetermined and its least squares optimal solution can be found with singular value decomposition (SVD).

The TM just described provides the registration between two given images, but cannot align arbitrary two images captured in the field where CPs are not available in real time. For this we expand the process to support image pairs of arbitrarily distant objects via pre-calibrated registration. The base of this approach is the Distance Dependent Transformation Matrix (DDTM) ($TM(D)$)—A TM that depends on distance D . We construct this matrix function as follows

- (i) Capture a scene with varying distances between the sensors and the CPs,
- (ii) Calculate the TM for each scene with its corresponding distance,
- (iii) Construct a collection of all the TM calculated (Figure 2),
- (iv) Collect the corresponding values from each element of each matrix, use them as samples of a distance-dependent function, and perform a regression R with these samples to obtain a functional representation of the element

$$DDTM(D)_{(m,n)} = R(TM(D_1)_{(m,n)}, TM(D_2)_{(m,n)}, \dots, TM(D_z)_{(m,n)})
 \tag{4}$$

where m and n are the row and column indices of the TM respectively,

- (v) Collect all functions to construct the DDTM.

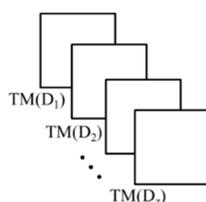


Figure 2. TM collection.

Using Equation (1) while defining the TM component as the newly defined DDTM (Equation (5)), each new pixel's coordinates will be calculated according to its distance D (which is assumed measured).

$$DDTM(D) = \begin{bmatrix} DDTM(D)_{(0,0)} & DDTM(D)_{(0,1)} & DDTM(D)_{(0,2)} \\ DDTM(D)_{(1,0)} & DDTM(D)_{(1,1)} & DDTM(D)_{(1,2)} \\ DDTM(D)_{(2,0)} & DDTM(D)_{(2,1)} & DDTM(D)_{(2,2)} \end{bmatrix} \quad (5)$$

The above equation is the core of a very compact registration of thermal and RGB camera images depending on the distance, which will be in our case measured with a laser scanner.

2.2. Target Rail and Sensor Position

The experimental system included two main elements: sensor array (Figure 3a) and a circular rail with mobile rail-cart (Figure 3b). The sensor array contained three sensors, including a RGB camera (Flea2 FL2-08S2C with resolution of 768×1032 and 45° wide angle lens, acquisition frequency 30 FPR), a thermal camera (Flir T425 with resolution of 240×320 and 45° IR wide angle lens, acquisition frequency 9 FPR), and a laser scanner (Sick LMS111 with scanning angle of 270° and resolution of 0.5° , scanning frequency 50 Hz). The sensors were mounted to the sprayer disabling any relative movement among sensors during all experiments. The second experimental element is a circular rail (width 6.4 m/length 2.4 m) with a mobile rail-cart (Figure 3b) able to move along the rail line. The rail-cart position/velocity can be controlled manually or by using an AC electric motor with frequency inverter regulation. Several types of targets can be fixed on top of the rail-cart such as artificial targets or real-world target (live trees). Data from all three sensors was acquired simultaneously using a personal computer and stored for post analysis.

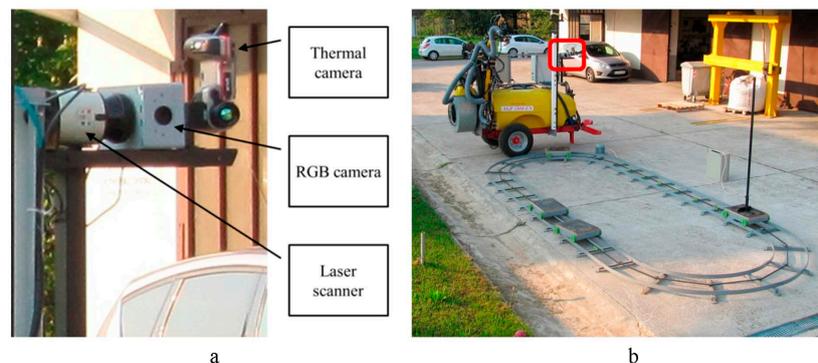


Figure 3. Selective sprayer. (a) sensors array; (b) overall sprayer with rail (red square, sprayer sensors).

2.3. Artificial Control Points

To construct the DDTM, the collection of TMs that register a preconfigured pair of sensors had to be derived from experimental data. To do so, a unique type of artificial control point (ACP) was designed with the goal to be easily detected by the two sensors, the RGB camera and the thermal camera. Each of our ACPs is composed of two elements, a colored disk (120 mm diameter) with 10 W incandescent light bulb fixed at the disk center (Figure 4a,c). The disks were colored differently to provide better visualization. While not used here, in future applications these different colors can be used to associate

unique identities to the markers in order to facilitate more elaborate inference about the spatial organization of the ACPs.

In order to calculate a reliable DDTM in a real world scene, eight ACPs were mounted to a flat plate (1×1 m) as shown in Figure 4b. The plate dimensions and the ACPs arrangement within the plate were chosen such that at the minimum operational distance between the plate and the sensors array (170 cm) the ACPs will cover most of the image area and provide maximal sensitivity in their position. In order to evaluate the DDTM, the ACPs flat plate was mounted to the rail-cart with the ability to move while the plate remains perpendicular to the sensors array (Figure 4d).

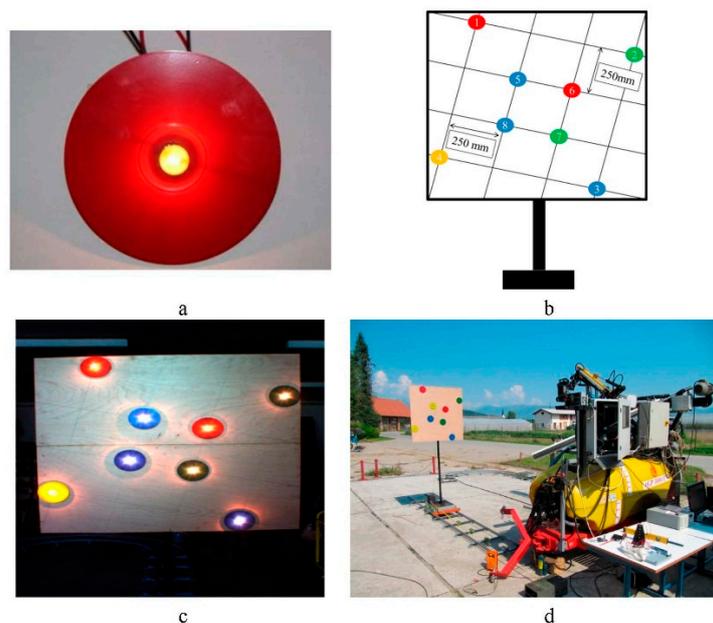


Figure 4. ACPs (a) single target with an incandescent light bulb in the middle; (b) ACPs setup; (c) ACPs mounting plate; (d) DDTM evaluation setup.

The colored disk detection algorithm was designed to detect and localize the ACPs in the color image and is based on commonly used machine vision procedures. The algorithm was implemented using Matlab software and Matlab image processing toolbox and was based on color thresholding (according to the specific disk). Figure 5 shows the detection of three blue disks in the image, but the algorithm is easily modified according to the desired target color. More specifically, the algorithm works as follows:

- (i) Capture input RGB image (Figure 5a),
- (ii) Convert the RGB image to HSV representation and isolate the hue and saturation channels (Figure 5b,d where (b) is the Hue and (d) is the saturation),
- (iii) Threshold the hue and the saturation (thresholds were set according to the detected color) (Figure 5c,e),
- (iv) Merge (logical OR) the resulted binary images (step iii, Figure 5f),
- (v) Isolate the ACPs plate using the RGB image (Figure 5g) (the ACPs plate is brighter relative to the environment and by applying threshold on the R,G,B channels the ACPs plate can be identified and isolate),
- (vi) Unite the binary ACP plate (Figure 5g) with the binary image resulted in step iv (Figure 5f),

- (vii) Remove small clusters (<500) of pixels that considered as noise (using Matlab command *bwareaopen*) (Figure 5i),
- (viii) Fill holes in the image using morphological operations (using Matlab command *imfill*) (Figure 5j),
- (ix) Apply erosion followed by dilation using disk mask ($r = 15$ pixel) (Figure 5k). Using the disk shape mask, this step helps to remove small, non-ellipse shapes,
- (x) Filter remaining identification errors by searching saturated pixels in the shape center (expected due to the light bulb) (Figure 5l).

The detection of ACPs in the thermal images, *i.e.*, the detection of the light bulbs, followed the following steps:

- (i) Capture a thermal image,
- (ii) Threshold the image for high temperature values,
- (iii) Remove small clusters (<10) of pixels that considered as noise (using Matlab command *bwareaopen*),
- (iv) Apply erosion followed by dilation using disk mask ($r = 5$). Using the disk shape mask, this step helps to remove small, non-ellipse shapes.

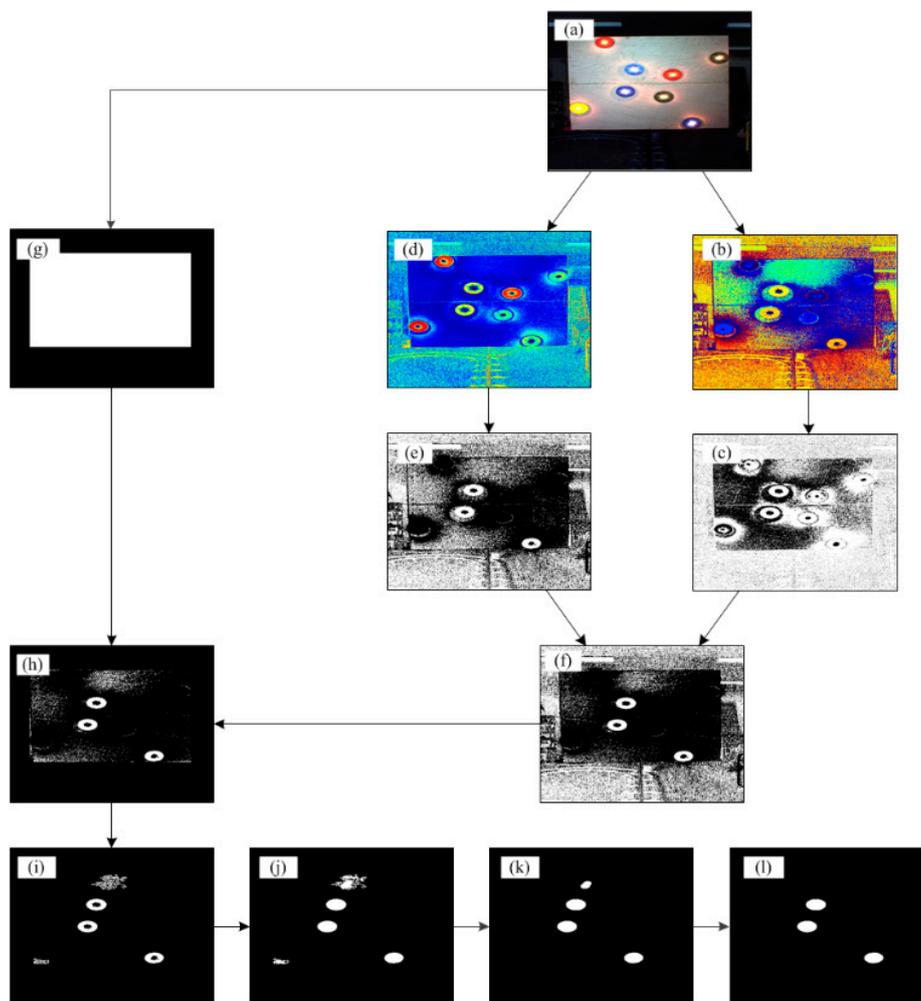


Figure 5. Blue colored disk detection algorithm.

The ACP position in the visual and thermal image was calculated as the center of mass of each ACP position (Figure 6).

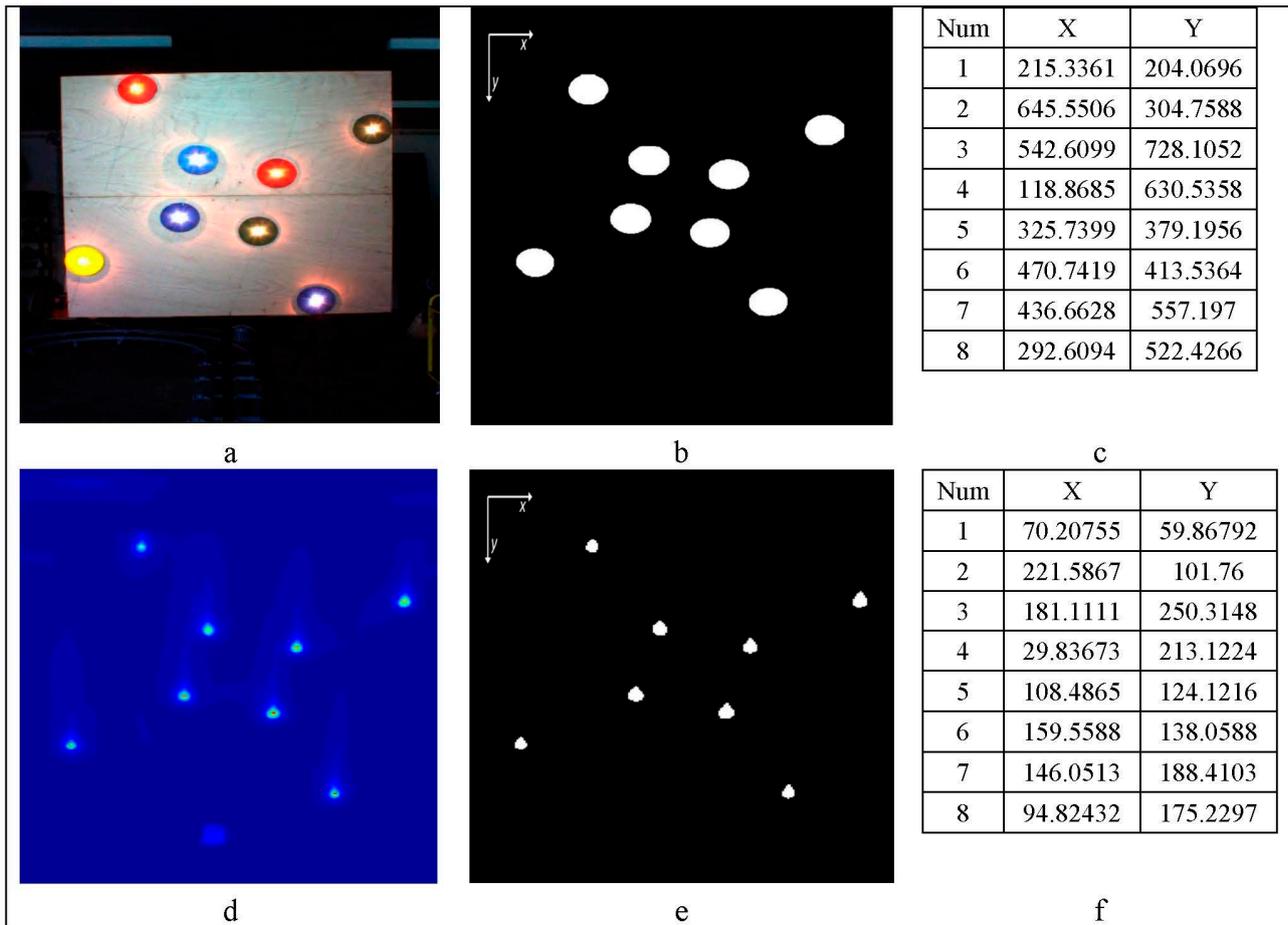


Figure 6. (a) visual image; (b) RGB ACPs positions, (c) RGB ACPs position; (d) thermal image; (e) thermal ACPs position; (f) ACPs position; (c,f) are sorted according to Figure 4b.

2.4. ACP Analysis

To validate the ACPs detection algorithms, we compared their detected locations with ground truth position marked by a human operator on a set of 192 points (24 random images, eight points in each image). The distance d_{RGB} pixels, between the automatically and manually marked ACPs was considered as the detection error and calculated by the standard Euclidean formula

$$\begin{aligned}
 d_{RGB} &= \sqrt{(x_{RGB} - x_{MRGB})^2 - (y_{RGB} - y_{MRGB})^2} \\
 d_{Thermal} &= \sqrt{(x_{Thermal} - x_{MThermal})^2 - (y_{Thermal} - y_{MThermal})^2}
 \end{aligned}
 \tag{6}$$

where d_{RGB} is the distance in pixels, x_{RGB} and y_{RGB} are the automatically calculated horizontal and vertical coordinates of the visual image respectively, x_{MRGB} and y_{MRGB} are the manually marked horizontal and vertical coordinates of the visual image respectively, $x_{Thermal}$ and $y_{Thermal}$ are the automatically calculated horizontal and vertical coordinates of the thermal image respectively, $x_{MThermal}$ and $y_{MThermal}$ are the manually marked horizontal and vertical coordinates of the thermal image respectively. Since the visual and thermal cameras have different resolutions, a normalization correction was added according to Equation (7). Needless to say, these constants require adjustment for other cameras and lenses.

$$d_{RGB} = \sqrt{\left((x_{Thermal} - x_{MThermal}) \cdot \frac{1032}{320} \right)^2 + \left((y_{Thermal} - y_{MThermal}) \cdot \frac{768}{240} \right)^2} \quad (7)$$

Lens distortions play an important role in the registration procedure. Although distortion can be irregular or follow many patterns, the most commonly encountered distortions are approximately radially symmetric, arising from the symmetry of a photographic lens. This is true for both thermal and visual light lenses. However, results summarized later in Table 3 show that these were not very large and that in accordance h20 and h21 were close to zero.

Table 2 shows that the average error between the automatically and manually marked ACPs is 1.36 pixel and 0.87 pixel for the visual and the thermal image respectively. Comparison between the ACPs detection algorithms shows that the color disk detection algorithm is more accurate than the thermal detection algorithm (for this comparison, normalized thermal data is used).

Table 2. Error comparison between manual and automatically ACPs detection.

Sensor	Average	Standard Deviation
	pixel	pixel
RGB	1.36	0.88
Thermal	0.87	0.32
Thermal (normalized)	2.7	1.03

3. Experimental Evaluation

With the approach outlined above, experimental evaluation was performed in three steps. First, we constructed the DDTM using measured data. Second, we evaluated the constructed DDTM under easy/simple conditions. Evaluation of performance in real-world conditions is described in the third step which measures performance for different vibrations (pulling tractor, wind) and target registration of oblique objects.

3.1. Experiment 1: Estimation of the DDTM

The first experiment goal was to create the $DDTM(D)$. Note that this computation is needed for each multimodal sensory configuration and can be considered as its calibration. Doing so requires sampling of corresponding ACPs, solving their TMs from several distances and then performing regression on the numerical results (as described in Section 2.1).

In order to calculate the $DDTM(D)$ in our experiment, a set of 51 scenes were captured from each distance and was used for better estimation of the ACPs position. The 51 scenes were captured sequentially during time frame of ~5.5 s. The scenes were captured by varying the distances between the sensor array and the ACPs plate from 1700 to 3000 mm with intervals of 50 mm resulting in a total of 27 distances, 408 ACPs for each distance, and a total of 1377 images which contributed to the DDTM estimation. These distances were selected because they correspond well to the operational distances of the specific sprayer system we used for evaluation. Each captured scene includes a single visual image from the RGB camera, a single thermal image, and a single laser scan that includes the plate surface.

According to the suggested model in Section 2.1, a linear regression was used on each of the TM elements (Figures 7 and 8 and Table 3). The linear equations of h_{00} , h_{01} , h_{10} , h_{11} , h_{20} and h_{21} were approximated to a constant value (the second linear coefficient) since the slope coefficient was extremely small (Table 3), *i.e.*, only h_{02} and h_{21} were affected by the distance between the plate and the sensor array. The $DDTM(D)$ calculated for our particular sensor array was therefore presented as follows:

$$DDTM(D) = \begin{bmatrix} 0.3729 & -0.01408 & -0.012784 \cdot D + 16.110487 \\ 0.025381 & 0.363118 & -0.000845 \cdot D - 17.503239 \\ 0.000090 & -0.000014 & 1 \end{bmatrix} \quad (8)$$

where D mm is the distance between the sensors array and the target for which registration is performed.

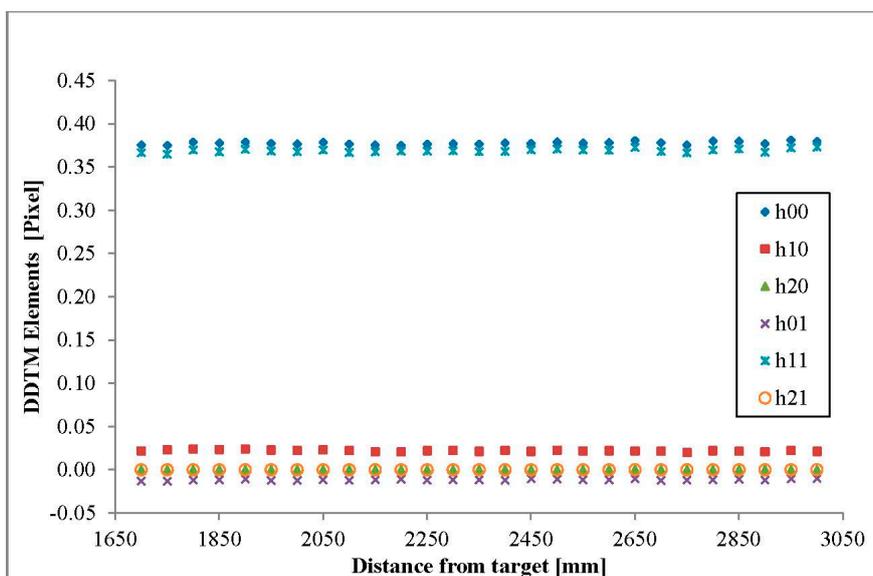


Figure 7. DDTM elements (h_{00} , h_{10} , h_{20} , h_{01} , h_{11} , h_{21}).

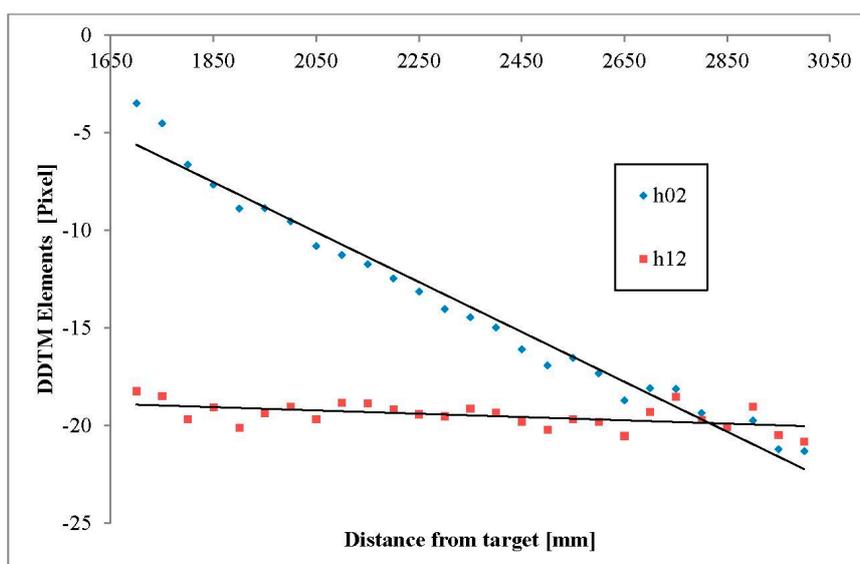


Figure 8. DDTM elements (h_{02} , h_{12}).

Table 3. DDTM regression summary for the sensor configuration used in our experiment.

$DDTM_{(m,n)}(D)$	Regression Result	Approximation as Constant
h_{00}	$f(D) = 2.15E-06 \cdot D + 3.72E-01$	0.3729
h_{01}	$f(D) = 1.06E-06 \cdot D - 1.41E-02$	-0.014085
h_{02}	$f(D) = -0.012784 \cdot D - 16.110487$	
h_{10}	$f(D) = -1.44E-06 \cdot D - 2.54E-02$	0.025381
h_{11}	$f(D) = 2.44E-06 \cdot D - 3.63E-01$	0.363118
h_{12}	$f(D) = -0.000845 \cdot D - 17.503239$	
h_{20}	$f(D) = -5.56E-09 \cdot D + 9.00E-05$	0.000090
h_{21}	$f(D) = 3.71E-09 \cdot D - 1.45E-05$	-0.000014

Equation (8) and the results summarized in Table 3 indicate that h_{20} and h_{21} are close to zero which suggests that the resulting DDTM is an affine TM type with the following shape:

$$\text{affine TM} = \begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix} \quad (9)$$

where A is a 2×2 matrix responsible for the rotation and deformation and t responsible for the translation [29]. A elements in Equation (9) can be decomposed and pose as:

$$A = R(\Theta)R(-\Phi)DR(\Phi) \quad (10)$$

where $R(\Theta)$ and $R(\Phi)$ are rotation by Θ and deformation by Φ , and D is a diagonal scaling matrix:

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

If the two cameras were perfectly aligned in parallel (horizontal and vertical) then Θ and Φ values would be zeros, resulting in:

$$A = D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad (11)$$

and the values of λ_1 and λ_2 would be the cameras resolution ratios,

$$\lambda_1 = \frac{320}{1032} = 0.31$$

$$\lambda_2 = \frac{240}{768} = 0.3125$$

But since our sensors also sustained minor misalignment setup (as expected in any field setup), the Θ and Φ are not zeros and the results of A are DDTM elements (h_{00} , h_{01} , h_{10} , h_{11}).

The translation elements (h_{02} and h_{12}) are linear as shown in the following:

Consider the rectified setup shown in Figure 9, where two cameras are horizontally aligned and the distance between their centers (a.k.a. the baseline) is d . The thermal and color cameras horizontal fields of view are α_1 and α_2 , respectively. The distance between the cameras and the target is D . In order to horizontally translate (t_H) the left pixel of the color camera (Figure 9 blue circle) to the corresponding left of the thermal camera, the following should be calculated:

$$\begin{aligned}
 t_H &= x_1 + d + x_2 \\
 x_1 &= D \cdot \tan\left(\frac{\alpha_1}{2}\right) \\
 x_2 &= D \cdot \tan\left(\frac{\alpha_2}{2}\right) \\
 t_H &= D \cdot \tan\left(\frac{\alpha_1}{2}\right) + d + D \cdot \tan\left(\frac{\alpha_2}{2}\right) \\
 t_H &= D \cdot \left(\tan\left(\frac{\alpha_1}{2}\right) + \tan\left(\frac{\alpha_2}{2}\right) \right) + d
 \end{aligned}$$

and since α_1 and α_2 are constants, t_H is linear. The vertical translation can be shown to be linear in the same calculation procedure.

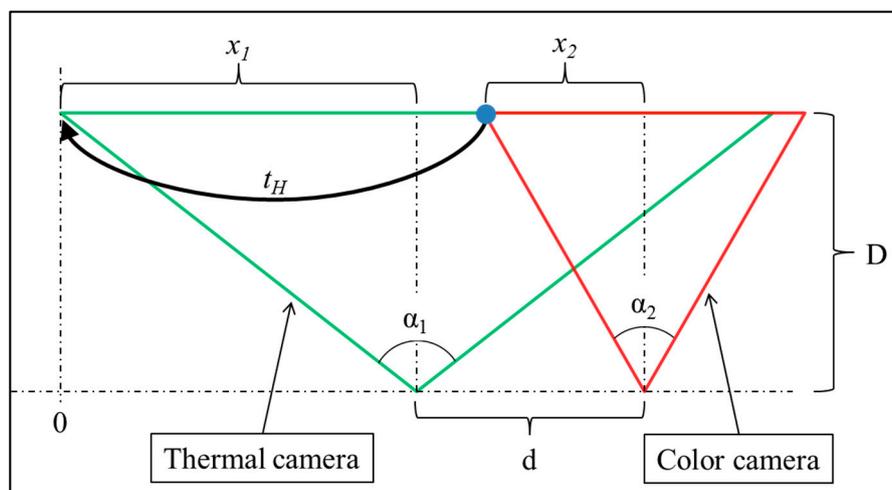


Figure 9. Horizontal translation example.

3.2. Experiment 2: DDTM Evaluation, Straight Vertical Plane

Once the DDTM was determined as shown above, we next evaluated it for registration of controlled measured data. In this experiment the ACPs will change their designation and be used as image registration validation points. The ACP RGB coordinate will be mapped to the thermal image. Ideally, by applying the registration, the mapped RGB coordinates should align with the thermal coordinates, while any difference will be considered as a registration error calculated using:

$$\begin{aligned}
 x_e &= (x'_c - x_T) \cdot \frac{559}{RP_D} \\
 y_e &= (y'_c - y_T) \cdot \frac{559}{RP_D} \\
 d_e &= \sqrt{x_e^2 + y_e^2}
 \end{aligned}$$

where x_e and y_e are the errors in x and y coordinates of the registered color image (x'_c, y'_c) to the thermal image (x_T, y_T), d_e is the distance error and RP_D is the pixel distance between the red CPs in the thermal image. The error is normalized to millimeter units by multiplying the registration error by the physical distance between CPs 1 and 6 in Figure 4b

$$\sqrt{250^2 + 500^2} = 559 \text{ (mm)}$$

divided by the pixel distance between CPs 1 and 6 detected in the thermal image.

To run the evaluation, we acquired new sets of images of the visual and thermal ACPs at each of the distances used above (*i.e.*, 1700 to 3000 mm with 50 mm intervals between). After acquisition, the visual ACPs were mapped to the thermal ACPs using the DDTM(D) instances. For each distance, 51 scenes were captured. The registration error was calculated for each CPs in the capture scene (51 scenes for each distance) for each of the distances (1700~3000).

Registration error evaluation (Figure 10) reveals that while considering the error in terms of pixels, the error remains close to constant value ($[0, -2]$ for x, $[-2, -3]$ for y and $[3, 5]$ for the distance error) but when considering the error in terms of physical distance mm the error increases with the distance from the target (D). The maximum registration error was no more than 3–5 pixels when using the DDTM(D).

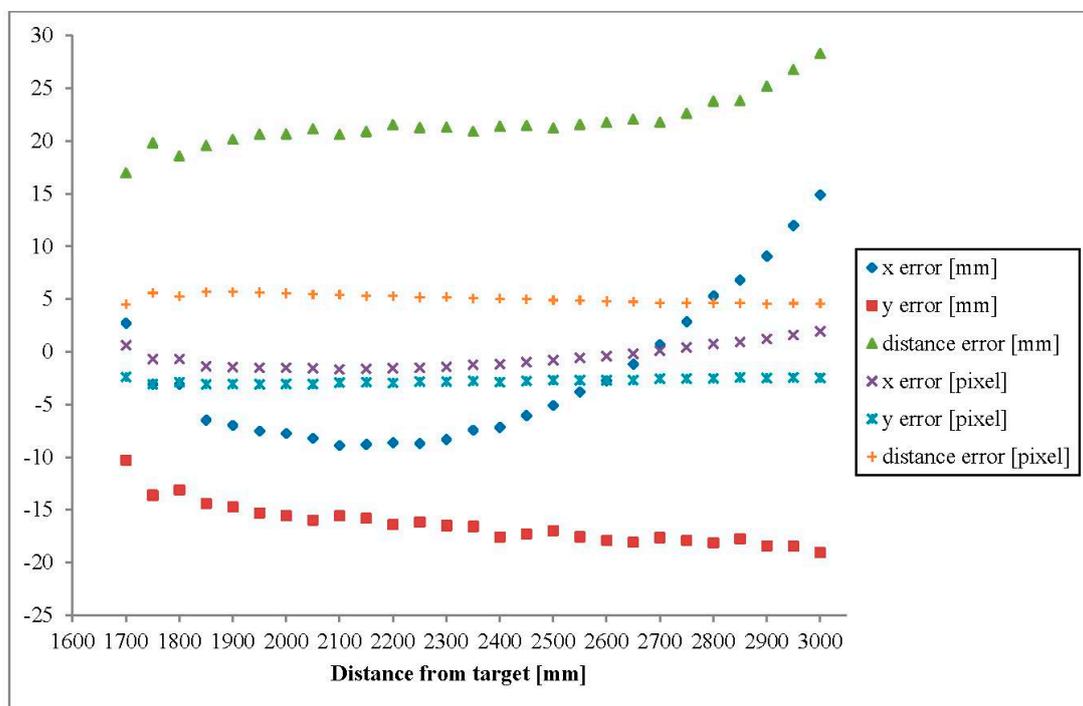


Figure 10. Registration errors while applying DDTM.

3.3. Experiment 3: DDTM Evaluation, Oblique ACPs Plane and Sensor Vibration

The objectives of the third experiment were to simulate more realistic sensory scenarios. We first examined performance when each image pixel is mapped according to the specific distance from the sensors. Second, we evaluated the registration performance while applying vibrations to the sensors unit, a condition imitating the vibrations that a pulling tractor would impose when driving and pulling the sprayer on uneven ground.

The experimental setup included two flat plates connected at 45° (Figure 11a). Four ACPs were attached to each of the plates with a total of 8 ACPs (Figure 11b). The oblique plates were mounted to the rail-cart with the ability to move while the plates remain perpendicular to the sensors array. Figure 11c shows sample of the visual image captured from the RGB camera.

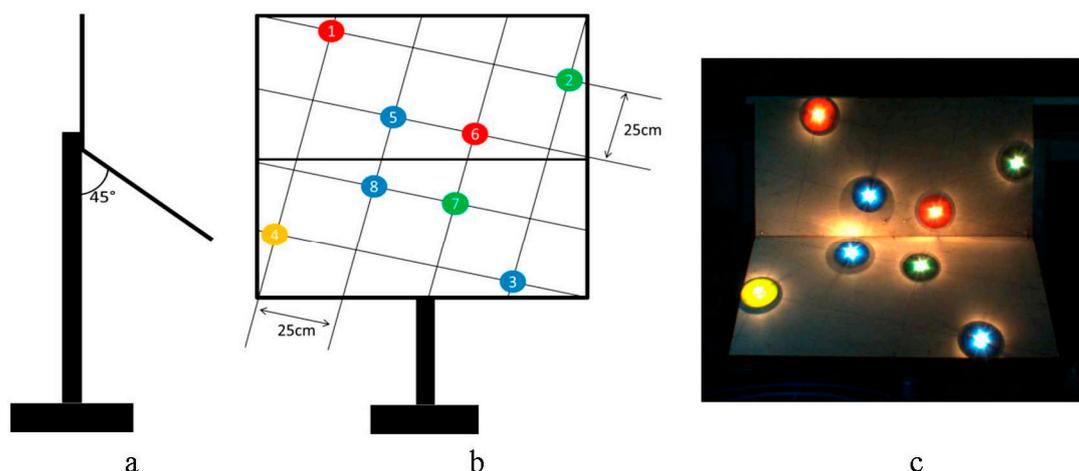


Figure 11. ACPs arrangement. (a) side view; (b) front view; (c) visual image.

During each experiment, the sprayer was stationary, while the rail cart with ACPs moved perpendicular to the sprayer and sensing unit. Sensing unit vibrations were in the Cartesian space. During the experiment, many Eigen—frequencies of the sprayer, half—filled reservoir, modifications to the sprayer *etc.* resulted in different sprayer (including sensing equipment) vibration levels for each coordinate of Cartesian space and for each vibration frequency setting. Due to a variety of vibration modes and few operating points the vibrations are expressed as total acceleration as shown in Table 4.

Table 4. Results of the third experiment.

Vibration Frequency Hz (Swing Arm Rotation Frequency Hz)	Acceleration g	Number of Captured Images	Vertical Plane CP1, CP2, CP5, CP6 pixel		Oblique Plane CP3, CP4, CP7, CP8 pixel	
			Mean Error	Standard Deviation	Mean Error	Standard Deviation
0	0	118	3.68	0.67	4.02	0.86
2.18	0.011	118	3.34	0.79	4.18	1.20
2.49	0.0577	118	3.29	0.72	4.42	1.21
2.80	0.0595	118	3.38	0.62	3.62	0.98
3.11	0.1117	118	3.07	0.70	3.81	1.04
3.42	0.1941	118	3.38	0.84	3.89	1.19
3.73	0.3467	118	4.07	1.33	4.19	1.51
4.04	0.3252	118	3.91	1.23	4.11	1.43
4.35	0.2433	118	3.45	0.87	3.72	1.32
4.66	0.2596	118	3.74	1.10	3.76	1.13
4.97	0.3575	118	3.55	0.96	4.20	1.33
5.28	0.4519	118	3.68	0.88	4.13	1.20

The vibrations were created using a frequency controlled AC electric motor with an eccentrically attached mass. The electric motor with its eccentrically attached mass was bolted to the sprayer's sensor array. Induced vibrations caused the entire sprayer to shake. During this experiment, the sprayer tank was filled with 500 liters of water, simulating a typical load. Vibrations were measured using an

accelerometer sensor (Phidgets 1043 with an acquisition frequency of 100 Hz) mounted rigidly on the sensing equipment support arm.

During the experiment, 12 electric motor rotational frequencies were evaluated, while the mass and its eccentric attached point remained fixed. The first frequency was set to zero (no vibration) and used as a reference to the vibration measurement. For each frequency, the rail-cart traveled from distance 1700 to 3000 mm and back to 1700 mm along the rail while the scene was captured continuously during the movement.

The mean registration error and the mean standard deviation were calculated for all test cases according to Table 4.

Comparison between the different vibrations shows that the vibration has no effect on the DDTM performance. This can be explained by the high frame rate of the sensors (30 and 9 frames per second for the color and thermal camera respectively).

4. Discussion and Conclusions

We have developed and presented a practical method for registration of multimodal sensory rigs for agricultural tasks. Our approach is based on pre-calibrating a distance-dependent TM between the sensors, and representing it in a compact way by regressing the distance-dependent coefficients as distance-dependent functions. In our case these dependencies ended up linear, but more elaborate DDTM may be obtained in more complicated situations.

Error evaluation of x and y coordinates show that while considering the registration error in the pixel space, the error remains relatively constant, but when considering the error in the Euclidean distance space the error increases with the distance between the sensors and the target.

The registration method suggested was developed for use with a one dimensional distance sensor (laser scanner). The method can be easily extended for a two dimensional distance sensor (e.g., Kinect sensor, Time of Flight camera) and gain higher accuracy from using the known distance of each image pixel.

While the presented approach was developed for an agricultural environment application, it can be applied to other applications that require registration of objects at varying distances. The method can be used to register images from all imaging sensors providing the sensors can detect common control points.

Acknowledgments

This research was partially supported by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative, by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering, and by the Frankel Fund, at the Computer Science Department, all at Ben-Gurion University of the Negev.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Kapach, K.; Barnea, E.; Mairon, R.; Edan, Y.; Ben-Shahar, O. Computer vision for fruit harvesting robots—state of the art and challenges ahead. *Int. J. Comput. Vis. Robot.* **2012**, *3*, 4–34.
2. Bulanon, D.; Burks, T.; Alchanatis, V. Image fusion of visible and thermal images for fruit detection. *Biosyst. Eng.* **2009**, *103*, 12–22.
3. Stajanko, D.; Lakota, M.; Hocevar, M. Application of Thermal Imaging for Visualization of High Density Orchards and Counting Apple Fruits. In Proceedings of the VII International Symposium on Modelling in Fruit Research and Orchard Management, Copenhagen, Denmark, 1 April 2006; pp. 211–215.
4. Jimenez, A.; Ceres, R.; Pons, J. A survey of computer vision methods for locating fruit on trees. *Trans. Am. Soc. Agric. Eng.* **2000**, *43*, 1911–1920.
5. Berenstein, R.; Shahar, O.B.; Shapiro, A.; Edan, Y. Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer. *Intell. Serv. Robot.* **2010**, *3*, 233–243.
6. Bulanon, D.; Burks, T.; Alchanatis, V. Study on temporal variation in citrus canopy using thermal imaging for citrus fruit detection. *Biosyst. Eng.* **2008**, *101*, 161–171.
7. Steward, B.L.; Tian, L.F.; Tang, L. Distance-based control system for machine vision-based selective spraying. *Trans. Am. Soc. Agric. Eng.* **2002**, *45*, 1255.
8. Zheng, J. Intelligent Pesticide Spraying Aims for Tree Target. In *Resour. Resource: Engineering & Technology for a Sustainable World*; ASABE: St. Joseph, MI, USA, 2005.
9. Shin, B.S.; Kim, S.H.; Park, J.U. In *Autonomous Agricultural Vehicle Using Overhead Guide*; Automation Technology for off-Road Equipment: Chicago, IL, USA, 2002.
10. Ogawa, Y.; Kondo, N.; Monta, M.; Shibusawa, S. Spraying robot for grape production. In *Springer Tracts In Advanced Robotics*; Springer: Berlin, Germany, 2006; pp. 539–548.
11. Kassay, L. Hungarian robotic apple harvester. In Proceedings of the International Summer Meeting of American Society of Agricultural Engineers, Charlotte, NC, USA, 21–24 June 1992.
12. Buemi, F.; Massa, M.; Sandini, G. Agrobot: A Robotic System for Greenhouse Operations. In Proceedings of the 4th Workshop on robotics in Agriculture, IARP, Toulouse, France, 30–31 October 1995; pp. 172–184.
13. Baeten, J.; Donné, K.; Boedrij, S.; Beckers, W.; Claesen, E. Autonomous fruit picking machine: A robotic apple harvester. *Field and Service Robotics*; Springer: Berlin, Germany, 2008; pp. 531–539.
14. Yamamoto, S.; Hayashi, S.; Yoshida, H.; Kobayashi, K. Development of a stationary robotic strawberry harvester with a picking mechanism that approaches the target fruit from below. *Jpn. Agric. Res. Q.* **2014**, *48*, 261–269.
15. Zitova, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000.
16. Lester, H.; Arridge, S.R. A survey of hierarchical non-linear medical image registration. *Pattern Recognit.* **1999**, *32*, 129–149.
17. Zheng, Z.; Wang, H.; Khwang Teoh, E. Analysis of gray level corner detection. *Pattern Recognit. Lett.* **1999**, *20*, 149–162.
18. Ton, J.; Jain, A.K. Registering landsat images by point matching. *IEEE Trans. Geosci. Remote Sens.* **1989**, *27*, 642–651.

19. Stockman, G.; Kopstein, S.; Benett, S. Matching images to models for registration and object detection via clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **1982**, *PAMI-4*, 229–241.
20. Pratt, W.K. Correlation techniques of image registration. *IEEE Trans. Aerosp. Electron. Syst.* **1974**, *AES-10*, 353–358.
21. Barnea, D.I.; Silverman, H.F. A class of algorithms for fast digital image registration. *IEEE Trans. Comput.* **1972**, *100*, 179–186.
22. Althof, R.J.; Wind, M.G.J.; Dobbins III, J.T. A rapid and automatic image registration algorithm with subpixel accuracy. *IEEE Trans. Med. Imaging* **1997**, *16*, 308–316.
23. Jarc, A.; Pers, J.; Rogelj, P.; Perse, M.; Kovacic, S. Texture features for affine registration of thermal (FLIR) and visible images. In Proceedings of the Computer Vision Winter Workshop, St. Lambrecht, Austria, 6–8 February 2007.
24. Davis, J.W.; Sharma, V. Background-subtraction using contour-based fusion of thermal and visible imagery. *Comput. Vis. Image Underst.* **2007**, *106*, 162–182.
25. Istenic, R.; Heric, D.; Ribaric, S.; Zazula, D. Thermal and visual image registration in hough parameter space. In Proceedings of the 14th International Workshop on Systems, Signals and Image Processing, Maribor, Yugoslavia, 27–30 June 2007; pp. 106–109.
26. Schaefer, G.; Tait, R.; Zhu, S.Y. Overlay of thermal and visual medical images using skin detection and image registration. In Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'06, New York, NY, USA, 30 August–3 September 2006; pp. 965–967.
27. Tait, R.J.; Schaefer, G.; Hopgood, A.A.; Zhu, S.Y. Efficient 3-d medical image registration using a distributed blackboard architecture. In Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'06, New York, NY, USA, 30 August–3 September 2006; pp. 3045–3048.
28. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer: Berlin, Germany, 2010.
29. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: New York, NY, USA, 2003.