Multi-modal Image Registration for Robotic Aerial Inspection using Mutual Information

Ehab Salahat†, Joe Coventry†, Andrew Thomson†† and Robert Mahony†
† Research School of Engineering, Australian National University, Canberra, ACT, Australia
†† CWP Renewables, Canberra, ACT, Australia

***Abstract*—In this work, we present a novel two-stage template registration approach based on mutual information to precisely register multi-modal sensor data of photo-voltaic modules as the first step in data processing for autonomous robotic inspection of renewable infrastructure. In stage 1 we construct a multi-resolution transform net that samples the global warp parameters space. A branch-and-bound algorithm is used to efficiently search the transforms net until a candidate solution is determined that lies in a local-basin of attraction for the true optimum. The algorithm then switches to a robust and efficient trust-region descent method. The proposed search determines the warp that aligns the PV module template to the sensor image to sub-pixel accuracy with quadratic convergence. By using a mutual-information cost criterion, the same template can be used to register multiple sensing modalities such as thermal, near-infra-red (NIR), hyper spectral, RGBD, etc., providing the operators with a suite of data that enables detailed diagnosis of faults and defects. The robustness of our approach is illustrated by performing mono-modal and multi-modal template registration for data collected from Australian solar farms. The experiments show good warp estimation as quantified by the intersection-of-union metric.**

I. Introduction

Solar energy is seen as a sustainable alternative source of energy for fossil fuels. The market for solar energy in both developed and developing countries is flourishing [1]. Energy scientists predict that there will an additional 30 TeraWatts of energy need by the middle of the 21st century and large-scale solar energy facilities will be a necessary part of the solution to meet this very challenging energy requirement [2]. Consequently, it is expected that there will be a large-scale deployment of solar power facilities, especially in the 66 sunbelt countries with more than 5 billion inhabitants representing 75% of the world's population.

Solar power facilities, such as photovoltaic (PV) and Concentrating Solar Power (CSP) facilities, suffer from many faults and defects that degrade their performance during their operational life. Detecting faults and defects requires regular inspection over physically large and distributed solar infrastructure. Presently, on-site visual inspection is performed by maintenance personnel on a regular basis to determine the deterioration level [3]. Manual inspection is costly and development of an autonomous robotic visual inspection system for solar infrastructure offers the potential to provide up-to-date and detailed information on the operational status of solar panels and mirrors, allowing operators to maximize value in the operation and maintenance of the facility. Moreover, solar power plants are often operated in inhospitable environments (e.g. deserts), and removing humans from such unfriendly environments is desirable [4]. Due to their increased availability, cost-effectiveness, small size, and maneuverability, Unmanned Aerial Vehicles (UAVs) are a natural candidate for robotic visual inspection of a large-scale infrastructure [5–7].

There are several trial studies using manually operated UAVs for survey and inspection of renewable infrastructure [8–11]. Simple image processing techniques were used to segment PV modules from high altitudes using infrared imagery [8]. The resulting segmentation results can be inaccurate due to non-homogeneous temperature distribution and Aghaei [11] concludes that such simple panel segmentation techniques are inappropriate for PV module inspection. Aghaei also used manually piloted UAVs for inspection [9] and developed algorithms to segment hotspots (indicating localised electrical or soiling defects on solar panels) using a thermal camera [12]. Although effective for identifying localized major faults, this approach is unable to detect more subtle faults, such as micro-cracks and incremental degradation or soiling of a panel, that can only be diagnosed using accurately registered images from multiple sensor modalities over extended periods (days, months or even years). The ability to autonomously acquire and accurately register sensor data from multiple sensing modalities (visible light, thermal, near infra-red, hyper-spectral, etc) is a key technical requirement for the effective long-term operation of robotic inspection for renewable energy infrastructure.

In this work, we present a novel two-stage mutual information ($\mathcal{MI}$) based template registration approach to register multi-modal sensor images of specific photo-voltaic modules. In the first stage, we create a pyramid of images and then build a multi-resolution transforms net [13] that coarsely samples a sub-space of the global parameter space of possible image warps. A branch-and-bound scheme [13] is used to efficiently search the transforms net starting with the low resolution images and progressively increasing resolution. The process terminates with a solution that is close, with a certain precision, to the optimum solution. In the second stage, we switch to a local search algorithm to obtain sub-pixel accuracy over the full parameter space. We use a robust trust-region algorithm combined with a line-search [14] method to obtain quadratic convergence with guaranteed descent. The warp update is carried out using the inverse-compositional Lucas-Kanade framework [15]. The overall efficiency of the template search is improved by selecting template pixels (required to construct the histograms used to compute the $\mathcal{MI}$ cost)

using an adaptive criterion based on Otsu's method [16]. Our approach can be used to register different sensing modalities (visual light, thermal, near infrared, hyperspectral) images to sub-pixel accuracy. The robustness of our approach is illustrated by performing mono-modal and multi-modal template registration experiments.

The remainder of this paper is organized as follows. Section II presents some preliminary information that is key for the sequel of the paper. Section III describes our proposed two-stage template registration methodology. Section IV presents experimental results that demonstrate the robustness of our approach. The paper's findings and contributions are summarized in section V.

II. Preliminaries

A. Mutual Information

Consider the question of registering a raw sensor image $\mathcal{I}$ to a given template image I^*. That is, we seek a parameterized warp function $w(\mathbf{x}, \mathbf{p})$ with parameters $\mathbf{p}$ and pixel coordinates $\mathbf{x}$ in some Euclidean space ($\mathbf{x} \in \mathbb{R}^2$) that overlays the template I^* on a raw image $\mathcal{I}$. Typically the raw image $\mathcal{I}$ will be larger than the template and only those pixels in the corresponding area should be included in the computation of the mutual information cost discussed in the sequel. Thus, for a given warp $w(\cdot)$, let I denote the sub-image of $\mathcal{I}$ corresponding to the area of the template I^* in shared (post warp) image coordinates. All pixel values of I and I^* are in the range $[0, 255]$. The mutual information framework models these intensity values as random variables sampled from continuous probability density function (PDF) $\mathcal{P}_I$ (resp. $\mathcal{P}_I^*$). Estimates p_I (resp. p_I^*) of $\mathcal{P}_I$ (resp. $\mathcal{P}_I^*$) can be computed by normalizing image intensity histograms using smooth B-splines

$$p_I(r, \mathbf{p}) = \frac{1}{N_\mathbf{x}} \sum_\mathbf{x} \beta_3(r - I(w(\mathbf{x}, \mathbf{p}))), \tag{1}$$

$$p_{I^*}(t) = \frac{1}{N_\mathbf{x}} \sum_\mathbf{x} \beta_3(t - I^*(\mathbf{x})), \tag{2}$$

where r (resp. t) defines the PDF bin of I (resp. I^*). Here $N_\mathbf{x}$ is the number of pixels. The PDF estimation formulation in (1)–(2) follows what is known as in-Parzen windowing [15]. The Parzen membership function in (1)–(2) is set to be a cubic B-Spline $\beta_3(\cdot)$ [17]. $\beta_3(\cdot)$ is defined as [17, 18]

$$\beta_3(s) = \begin{cases} (4 - 6|s|^2 + 3|s|^3)/6, & 0 \le |s| < 1 \\ (2 - |s|)^3/6, & 1 \le |s| \le 2 \\ 0, & \text{otherwise} \end{cases}. \tag{3}$$

The approach is best undertaken with only a few histogram bins. The resulting formulation is computationally efficient [15] and generates smooth cost functions. Note that (2) is independent of $\mathbf{p}$ as I^* and can be pre-computed. The joint distribution of I and I^* is estimated as

$$p_{II^*}(r, t, \mathbf{p}) = \frac{1}{N_\mathbf{x}} \sum_\mathbf{x} \beta_3(r - I(w(\mathbf{x}, \mathbf{p})))\beta_3(t - I^*(\mathbf{x})). \tag{4}$$

Shannon's marginal and joint entropies of I and I^* are computed from (1)–(4) to be [19]

$$\mathcal{H}(I) = \sum_r p_I(r, \mathbf{p}) \log(p_I(r, \mathbf{p})), \tag{5}$$

$$\mathcal{H}(I^*) = \sum_t p_I^*(t) \log(p_I^*(t)), \tag{6}$$

$$\mathcal{H}(I, I^*) = \sum_{r,t} p_{II^*}(r, t, \mathbf{p}) \log(p_{II^*}(r, t, \mathbf{p})). \tag{7}$$

Note that $\mathcal{H}(I^*)$ is only dependent on p_{I^*} and is a constant. While the entropy is a measure of the information content, the mutual information ($\mathcal{MI}$) is a measure of the shared information and statistical dependence [19]. The mutual information $\mathcal{MI}$ between I and I^* is evaluated from the marginal and joint entropies to be [19]

$$\begin{aligned} \mathcal{MI}(I, I^*) &= \mathcal{H}(I) + \mathcal{H}(I^*) - \mathcal{H}(I, I^*) \\ &= \sum_{r,t} p_{II^*}(r, t, \mathbf{p}) \log\left(\frac{p_{II^*}(r, t, \mathbf{p})}{p_I(r, \mathbf{p}) p_{I^*}(t)}\right). \end{aligned} \tag{8}$$

To illustrate how mutual information is effective in registering images it is necessary to consider the nature of the joint distribution p_{II^*}. For simplicity, consider the case of registering an image to itself, thus, the pixel intensity values will be identical when the image is correctly registered. Given that I and I^* have the same number of pixels ($N_\mathbf{x}$), the joint intensities of I and I^* are a list of 2-tuples

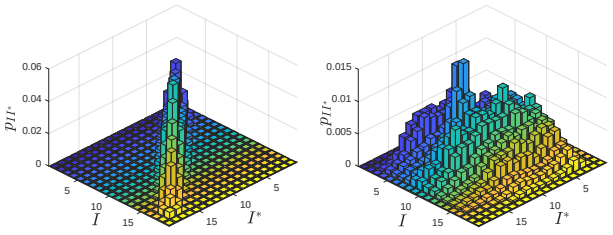
$$\{(I(w(\mathbf{x}_1, \mathbf{p})), I^*(\mathbf{x}_1)), ..., (I(w(\mathbf{x}_{N_\mathbf{x}}, \mathbf{p})), I^*(\mathbf{x}_{N_\mathbf{x}}))\}$$

that are generated by the correspondence of the shared (post-warp) pixel coordinates. Each pair will contribute to the joint histogram bin (r, t) of $p_{II^*}(r, t, \mathbf{p})$ only when $r = I(w_\mathbf{p}(\mathbf{x}))$ and $t = I^*(\mathbf{x})$. With perfect template alignment, the two intensity values in each 2-tuple in the list will be identical, resulting in a diagonal joint distribution (Fig. 1(a)). If the images are out of alignment the joint intensity values will no longer correspond, and the distribution becomes spread across the full intensity space (Fig. 1(a)) as the pixel locations for a given intensity in one image correspond to a range of different intensity values in the other image. The joint entropy measures the spread of the probability distribution; a concentrated probability distribution has high entropy, while a spread distribution has low entropy. Maximizing mutual information corresponds to identifying the warp that best aligns the images in the sense that it minimizes the spread and maximizes the concentration of the joint probability distribution. That is, for a given intensity value in one image, the corresponding pixels in the other image also correspond (or at least approximately correspond) to a single intensity value.

B. Image Pyramids

Image pyramid construction is a technique that represents an image using multi-resolution copies. Construction of an image pyramid involves repeated application of low-pass filtering followed by a REDUCE operation [17].

Denote the level of the image pyramid by $\ell = 1, ..., \ell_{scales}$ with $\ell = 1$ representing the lowest resolution and $\ell = \ell_{max}$



(a) (b)

Fig. 1: Comparison of two p_{II^*} for (a) perfectly aligned and (b) misaligned templates.

representing the original full resolution image resolution. Let $\eta \in (0,1)$ denote the scale reduction. Define σ_η to be

$$\sigma_\eta = \sigma_0 \sqrt{\eta^{-2} - 1}, \tag{9}$$

where $\sigma_0 = 0.6$ [20]. Then the image pyramid is built by iterating the following equation [21]:

$$\mathcal{I}^{\ell-1} = \mathcal{BI}_\eta \left(\mathcal{G}(\sigma_\eta) * \mathcal{I}^\ell(\mathbf{x}) \right), \tag{10}$$

where $\mathcal{BI}_\eta(\cdot)$ denotes down-sampling using bilinear interpolation with pixel sample distance $1/\eta$, '$*$' defines the convolution operation, $\mathcal{G}(\cdot)$ is a Gaussian filter with a kernel σ_η. Let $\Delta_\ell^{\mathcal{I}} = (\mathcal{I}^1, \ldots, \mathcal{I}^\ell)$ and $\Delta_\ell^{I^*} = (I^{*1}, \ldots, I^{*\ell})$ denote the image pyramids constructed for $\mathcal{I}$ and I^*.

C. Warp Parametrization

We use the maneuverability of the UAV system to position the drone camera as close as possible to the desired image registration as possible. The drone is servo-controlled until the image is approximately fronto-parallel to the photo-voltaic panel. It is not possible to servo control the drone sufficiently accurately to remove all relative rotation from the resulting image capture, however, the positioning accuracy is sufficient that we can restrict our attention to a subspace of 2D translation and scaling warps for the global search. This is a critical assumption, since it significantly reduces the dimension of the search space for the image transforms net.

The 2D translation and scaling warp is parametrized as

$$\mathcal{W}_s(\mathbf{p}_s) = \begin{bmatrix} s+1 & 0 & x \\ 0 & s+1 & y \\ 0 & 0 & 1 \end{bmatrix}, \tag{11}$$

where $\mathbf{p}_s = \begin{bmatrix} s & x & y \end{bmatrix}$, s is a scaling factor, while x and y are the translation (in pixels) in the respective direction.

For accurate image registration, it is necessary to also model perspective changes. We use general planar homographies for the warp parametrization for the local search algorithm. A homography matrix is given as

$$\mathcal{W}_h(\mathbf{p}_h) = \begin{bmatrix} p_1+1 & p_4 & p_7 \\ p_2 & p_5+1 & p_8 \\ p_3 & p_6 & 1 \end{bmatrix}, \tag{12}$$

where $\mathbf{p}_h = \begin{bmatrix} p_1 & p_2 & \ldots & p_8 \end{bmatrix}$.

TABLE I: Transforms Net for (11).

Parameter	Range	Step Size
x (pixels)	$[-(N_I - N_I^*)/2, +(N_I - N_I^*)/2]$	δ_x
y (pixels)	$[-(M_I - M_I^*)/2, +(M_I - M_I^*)/2]$	δ_y
s	$[M_I^*/M_I - 1, M_I/M_I^* - 1]$	δ_s

TABLE II: Branch-and-Bound transforms Net parameters for the (11).

Parameter	Range	Step Size
$x^{\ell+1}$ (pixels)	$[x_{opt}^\ell - \delta_x^\ell, x_{opt}^\ell + \delta_x^\ell]$	$\delta_x^{\ell+1} = \delta_x^\ell/2$
$y^{\ell+1}$ (pixels)	$[y_{opt}^\ell - \delta_y^\ell, y_{opt}^\ell + \delta_y^\ell]$	$\delta_y^{\ell+1} = \delta_y^\ell/2$
$s^{\ell+1}$	$[s_{opt}^\ell - \delta_s^\ell, s_{opt}^\ell + \delta_s^\ell]$	$\delta_s^{\ell+1} = \delta_s^\ell/2$

III. Registration Methodology

Consider a floating template I^* with dimensions $(M_{I^*} \times N_{I^*})$. The goal is to accurately determine the warp parameters $\mathbf{p}$ that will displace I^* to the closest match in $\mathcal{I}$ (cf. §. We propose an algorithm with two stages: Inspired by [13], the first stage performs a *global search* over a reduced subspace of warp parameters using a multi-resolution *transform net* to roughly determine the best match. The algorithm then switches to the second stage and performs a trust region search to determine the local sub-pixel warp parameters, using the output from the global search as an initialization point.

A. Global Search

Two image pyramids, $\Delta_{\mathcal{I}}^\ell$ and $\Delta_{I^*}^\ell$, are created for $\mathcal{I}$ and I^* with $\eta = 0.5$ (cf. §II-B). A multi-resolution *transforms net* is created for $\Delta_{\mathcal{I}}^\ell$. A *transforms net* is a set of transforms that discretize the warp parameters space. We propose to define the transforms net iteratively as the different levels of the pyramid are considered.

Consider the lowest level of the pyramid. A transform net of the warp parameters (11) is constructed as shown in Table I. The search proceeds by evaluating the $\mathcal{MI}$ at each transform net location in the lowest resolution image and selecting the maximum to find $\mathbf{p}_{opt}^1$:

$$\mathbf{p}_{opt}^1 = \underset{\mathbf{p}}{\operatorname{argmax}}\, \mathcal{MI}(\Delta_1^I, \Delta_1^{I^*}). \tag{13}$$

The optimal warp parameter $\mathbf{p}_{\text{opt}}^1$ is branched to the next level of the multi-resolution transforms net by defining

$$\mathbf{p}_{\text{ref}}^2 = \begin{bmatrix} s_{\text{opt}}^1 & x_{\text{opt}}^1/\eta & y_{\text{opt}}^1/\eta \end{bmatrix}. \tag{14}$$

Applying the branch and bound principle, we restrict the search in the higher resolution pyramid to the neighbourhood of the inherited optimum $\mathbf{p}_{\text{ref}}^2$. The new parameters' boundaries are given in Table II. This guarantees that $\mathbf{p}_{opt}^2$ will be at least as good as $\mathbf{p}_{opt}^1$.

The process is iterated for each level ℓ of the image pyramid until either the search terminates when $\ell = \ell_{max}$ or when the change in the parameters is smaller than a specified threshold.

While the computational complexity of the transforms net for (11) is relatively low, the search performance can be improved by selecting only template pixels with high information content in the computations. One approach [13] is to use a random pixel selection approach. Dame *et al.* [22] proposed to

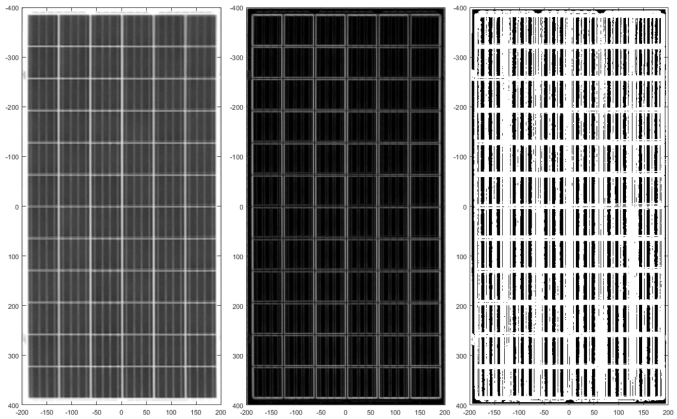


Fig. 2: (Left) A PV module template, its (middle) gradient image and (right) Otsu's thresholding result.

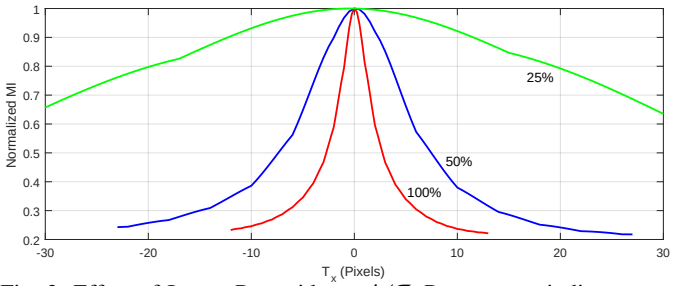


Fig. 3: Effect of Image Pyramids on $\mathcal{MI}$. Percentages indicate new resolution w.r.t. the original resolution.

use only pixels with strong gradient values, fixing a threshold τ and choosing a selection of pixels S_τ according to

$$S_\tau = \{\mathbf{x} \mid \|\nabla I^*(\mathbf{x})\| > \tau.\} \tag{15}$$

for $\nabla I^*(\mathbf{x})$ the image gradient at a pixel $\mathbf{x}$. While this approach is relatively simple to compute, it has, however, the disadvantage of a hard-coded threshold value. We propose an improvement based on using Otsu's method [16] to adaptively find the optimal threshold value that separates strong gradient pixels from the weak. An illustration of (15) with an adaptively defined threshold using Otsu's method is shown in Fig. 2. White pixels mask the pixels that will be involved in the computations that are related to evaluating of $\mathcal{MI}$ throughout the template registration process.

The multi-resolution global search approach reduces the number of transforms that need to be evaluated and simplifies the evaluation of mutual information, as the number of pixels involved is less at lower image resolutions. Note that as the resolution decreases, the information in the scenes becomes blurred, flattening the cost function as shown in Fig. 3. If the image pyramid is too deep, then the maximum of the cost at the lowest level of the pyramid can become ill constrained and the branch and bound can fail to converge to the basin of attraction of the optimum global solution.

B. Local Trust Region Search

In order to achieve a very precise estimation of the warp parameters that results in the best match, the algorithm switches from the pyramidal branch-and-bound search to an MI local trust region optimizer. The optimizer iteratively finds the

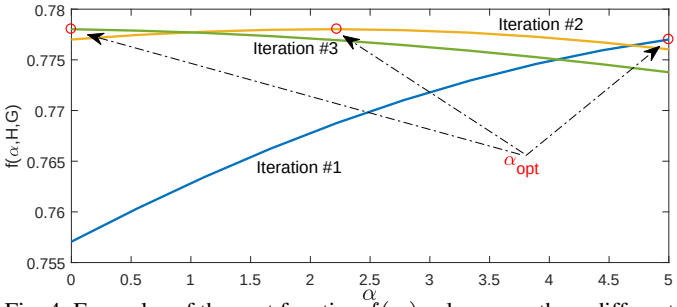


Fig. 4: Examples of the cost function $f(-)$ values over three different iterations and the optimal gain using Brent's algorithm.

incremental refinement parameters $\mathbf{\Delta p}$ that maximizes $\mathcal{MI}$ given $\mathbf{p}$, until a stopping criterion is met. That is:

$$\mathbf{\Delta p}^k = \underset{\mathbf{\Delta p}}{\operatorname{argmax}}\, \mathcal{MI}(I(w(\mathbf{x}, \mathbf{p}^k)), I^*(w(\mathbf{x}, \mathbf{\Delta p}))). \tag{16}$$

where k is the iteration number. When $k = 1$, $\mathbf{\Delta p} = 0$ and $\mathbf{p}$ is given by the global search results from section III-A. When $k \neq 1$, we use the iterative inverse compositional parameter refinement to find the new warp parameters due to its computational efficiency (see [15]). That is:

$$w(\mathbf{x}, \mathbf{p}^{k+1}) \leftarrow w(w^{-1}(\mathbf{x}, \mathbf{\Delta p}^k), \mathbf{p}^k). \tag{17}$$

Obviously, the update rule requires finding $\mathbf{p}^{k+1}$ by inversely composing $\mathbf{\Delta p^k}$ with $\mathbf{p^k}$. That is, one needs to find:

$$\mathcal{W}(\mathbf{p}^{k+1}) = \mathcal{W}(\mathbf{p}^k) \times \mathcal{W}^{-1}(\mathbf{\Delta p}^k), \tag{18}$$

where $\times$ refers to matrix multiplication. The homography parametrization $\mathcal{W}(\mathbf{p})$ is given in (12).

As can be seen from (16)–(18), the entire parameters' update relies on $\mathbf{\Delta p}$. For this reason, an accurate estimation of $\mathbf{\Delta p}$ is required. Moreover, one needs to reach the solution with the minimum number of iterations. For these two reasons, we combine the line-search step tuning [14] to find the α_{opt}, the optimal gain value for $\mathbf{\Delta p}$ that maximizes MI in the direction of the cost function's gradients. Figure 4 illustrates α_{opt} over the first few iterations for a typical example.

The robust trust-region algorithm is used. One solves

$$\mathbf{\Delta p}^k = -\alpha_{opt} \left(\mathbf{H} + \alpha_2 \text{diag}(\mathbf{H})\right)^{-1} \mathbf{G}^\top, \tag{19}$$

where $\mathbf{G}$ and $\mathbf{H}$ are the gradient and Hessian matrices of the $\mathcal{MI}$ with respect to $\mathbf{\Delta p}$ and are respectively given as [15]:

$$\mathbf{G} = \frac{\partial \mathcal{MI}(w(I^*, \Delta\mathbf{p}), w(I, \mathbf{p}))}{\partial \Delta \mathbf{p}}, \tag{20}$$

$$\mathbf{H} = \frac{\partial^2 \mathcal{MI}(w(I^*, \Delta\mathbf{p}), w(I, \mathbf{p}))}{\partial \Delta \mathbf{p}^2}. \tag{21}$$

Using the derivative chain rule, $\mathbf{G}$ and $\mathbf{H}$ can be written as:

$$\mathbf{G} = \sum_{r,t} \frac{\partial p_{II^*}}{\partial \Delta \mathbf{p}} \left(1 + \log\left(\frac{p_{II^*}}{p_{I^*}}\right)\right), \tag{22}$$

$$\mathbf{H} = \sum_{r,t} \left\{ \frac{\partial p_{II^*}}{\partial \Delta \mathbf{p}}^T \frac{\partial p_{II^*}}{\partial \Delta \mathbf{p}} \left(\frac{1}{p_{II^*}} - \frac{1}{p_{I^*}}\right) + \frac{\partial^2 p_{II^*}}{\partial \Delta \mathbf{p}^2} \left(1 + \log\left(\frac{p_{II^*}}{p_I^*}\right)\right) \right\}, \tag{23}$$

where p_{I^*} and p_{II^*} are given in (1) and (4). Note that p_{I^*} can be pre-computed as the template I^* is fixed over iterations.

In addition, given that template is close to the actual solution due to the prior global parameters' screening, we can safely assume (and approximate) that $I \approx I^*$ and the cost function is locally quadratic. Hence, $\mathbf{H}$ can be also approximated and pre-computed.

The first and second order derivatives ($\frac{\partial p_{II^*}}{\partial \Delta \mathbf{p}}$ and $\frac{\partial^2 p_{II^*}}{\partial \Delta \mathbf{p}^2}$) are given respectively as [22]:

$$\frac{\partial p_{II^*}}{\partial \Delta \mathbf{p}} = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi(r - I(w(\mathbf{x}, \mathbf{p}))) \frac{\partial \phi(t - I^*(w(\mathbf{x}, \Delta \mathbf{p})))}{\partial \Delta \mathbf{p}}, \tag{24}$$

$$\frac{\partial^2 p_{II^*}}{\partial \Delta \mathbf{p}^2} = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi(r - I(w(\mathbf{x}, \mathbf{p}))) \frac{\partial^2 \phi(t - I^*(w(\mathbf{x}, \Delta \mathbf{p})))}{\partial \Delta \mathbf{p}^2}, \tag{25}$$

where we recall here that $\phi(-)$ represents the cubic B-spline function as defined in (3). To compute $\partial \phi(-)/\partial \Delta \mathbf{p}$ and $\partial^2 \phi(-)/\partial \Delta \mathbf{p}^2$, we use the chain rule again to obtain:

$$\frac{\partial \phi(t - I^*(w(\mathbf{x}, \Delta \mathbf{p})))}{\partial \Delta \mathbf{p}} = -\frac{\partial \phi}{\partial t} \frac{\partial I^*}{\partial \Delta \mathbf{p}}, \tag{26}$$

$$\frac{\partial^2 \phi(t - I^*(w(\mathbf{x}, \Delta \mathbf{p})))}{\partial \Delta \mathbf{p}^2} = -\frac{\partial \phi}{\partial t} \frac{\partial^2 I^*}{\partial \Delta \mathbf{p}^2} + \frac{\partial^2 \phi}{\partial t^2} \frac{\partial I^*}{\partial \Delta \mathbf{p}}^T \frac{\partial I^*}{\partial \Delta \mathbf{p}}, \tag{27}$$

where $\partial \phi/\partial t$ and $\partial^2 \phi/\partial t^2$ are the first and second order derivatives of $\beta_3(t - I^*(w(\mathbf{x}, \Delta \mathbf{p})))$.

The next expressions needed are $\partial I^*/\partial \Delta \mathbf{p}$ and $\partial^2 I^*/\partial \Delta \mathbf{p}^2$ which can be similarly computed using the chain rule again. These are given as [22]:

$$\frac{\partial I^*}{\partial \Delta \mathbf{p}} = \nabla I^* \frac{\partial w(\mathbf{x}, \mathbf{p})}{\partial \Delta \mathbf{p}}, \tag{28}$$

$$\begin{aligned} \frac{\partial^2 I^*}{\partial \Delta \mathbf{p}^2} = & \frac{\partial w(\mathbf{x}, \mathbf{p})}{\partial \Delta \mathbf{p}}^T \nabla^2 I^* \frac{\partial w(\mathbf{x}, \mathbf{p})}{\partial \Delta \mathbf{p}} \\ & + \nabla_x I^* \frac{\partial^2 w_x}{\partial \Delta \mathbf{p}^2} + \nabla_y I^* \frac{\partial^2 w_y}{\partial \Delta \mathbf{p}^2}. \end{aligned} \tag{29}$$

where $\nabla I^* = \begin{bmatrix} \nabla_x I^* & \nabla_y I^* \end{bmatrix}^\top$, whereas $\nabla_x I^*$ and $\nabla_y I^*$ are the directional gradient images. Likewise, $\nabla^2 I^* = \begin{bmatrix} \nabla_{xx} I^* & \nabla_{xy} I^* & \nabla_{yx} I^* & \nabla_{yy} I^* \end{bmatrix}^\top$ are the 2nd order directional gradient images. Moreover, $\partial w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}$ and $\partial^2 w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}^2$ are the Jacobian and the Hessian of the warp.

The motivation of using the inverse compositional approach can then be seen by observing that (26)–(29) are pre-computable for a fixed template, improving the overall efficiency of the presented method.

The remaining expressions to complete the derivations are $\partial w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}$ and $\partial^2 w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}^2$. To find the Jacobian of a given warp $w(\mathbf{x}, \mathbf{p})$, $w(\mathbf{x}, \mathbf{p})$ is first rewritten as:

$$w(\mathbf{x}, \mathbf{p}) = \begin{bmatrix} w_x(\mathbf{x}, \mathbf{p}) \\ w_y(\mathbf{x}, \mathbf{p}) \end{bmatrix}, \tag{30}$$

where $w_x(\mathbf{x}, \mathbf{p})$ (resp. $w_y(\mathbf{x}, \mathbf{p})$) is the warp function for the x (resp. y) component given the pixel coordinates $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}$ and warp parameters $\mathbf{p}$. For a homography warp (12), (30) is written as:

$$\begin{bmatrix} w_x(\mathbf{x}, \mathbf{p}) \\ w_y(\mathbf{x}, \mathbf{p}) \end{bmatrix} = \frac{1}{p_3 x + p_6 y + 1} \begin{bmatrix} (1 + p_1)x + p_4 y + p_7 \\ p_2 x + (1 + p_5) y + p_8 \end{bmatrix}. \tag{31}$$

To this end, $\partial w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}$ can then be computed as:

$$\frac{\partial w(\mathbf{x}, \mathbf{p})}{\partial \mathbf{p}} = \begin{bmatrix} \partial w_x(\mathbf{x}, \mathbf{p})/\partial \mathbf{p} \\ \partial w_y(\mathbf{x}, \mathbf{p})/\partial \mathbf{p} \end{bmatrix}, \tag{32}$$

where again for homographies, (32) is given as:

$$\begin{bmatrix} \partial w_x(\mathbf{x}, \mathbf{p})/\partial \mathbf{p} \\ \partial w_y(\mathbf{x}, \mathbf{p})/\partial \mathbf{p} \end{bmatrix} = \begin{bmatrix} \frac{\partial w_x(\mathbf{x}, \mathbf{p})}{\partial p_1} & \frac{\partial w_x(\mathbf{x}, \mathbf{p})}{\partial p_2} & \dots & \frac{\partial w_x(\mathbf{x}, \mathbf{p})}{\partial p_8} \\ \frac{\partial w_y(\mathbf{y}, \mathbf{p})}{\partial p_1} & \frac{\partial w_y(\mathbf{y}, \mathbf{p})}{\partial p_2} & \dots & \frac{\partial w_y(\mathbf{y}, \mathbf{p})}{\partial p_8} \end{bmatrix}. \tag{33}$$

By setting $\mathbf{p} = 0$ in (33), $\partial w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}$ are found as:

$$\begin{bmatrix} \partial w_x(\mathbf{x}, \mathbf{p})/\partial \mathbf{p} \\ \partial w_y(\mathbf{x}, \mathbf{p})/\partial \mathbf{p} \end{bmatrix} = \begin{bmatrix} x & 0 & -x^2 & y & 0 & -xy & 1 & 0 \\ 0 & x & -xy & 0 & y & -y^2 & 0 & 1 \end{bmatrix}. \tag{34}$$

Following a similar procedure, $\partial^2 w(\mathbf{x}, \mathbf{p})/\partial \Delta \mathbf{p}^2$, one finds that

$$\frac{\partial^2 w_x(\mathbf{x}, \mathbf{p})}{\partial \mathbf{p}^2} = \begin{bmatrix} 0 & 0 & -x^2 & 0 & 0 & -xy & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -x^2 & 0 & 2x^3 & -xy & 0 & 2x^2 y & -x & 0 \\ 0 & 0 & -xy & 0 & 0 & -y^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -xy & 0 & 2x^2 y & -y^2 & 0 & 2xy^2 & -y & 0 \\ 0 & 0 & -x & 0 & 0 & -y & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{35}$$

$$\frac{\partial^2 w_y(\mathbf{x}, \mathbf{p})}{\partial \mathbf{p}^2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -x^2 & 0 & 0 & -xy & 0 & 0 \\ 0 & -x^2 & 2x^2 y & 0 & -xy & 2xy^2 & 0 & -x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -xy & 0 & 0 & -y^2 & 0 & 0 \\ 0 & -xy & 2xy^2 & 0 & -y^2 & 2y^3 & 0 & -y \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -x & 0 & 0 & -y & 0 & 0 \end{bmatrix}. \tag{36}$$

IV. Experimental Results

To validate the performance and robustness of the proposed algorithm, we collected a set of images from several solar farms near Canberra. In our flights, we used DJI Matrice 600, a powerful hex-copter that is particularly designed for applications such as professional photography and industrial inspections (Fig. 5). The Matrice 600 can remain in the air even after the failure of one or two of its motors during flight. This ensures safety of the drone, nearby assets and most importantly people who may be working in the vicinity. It also has triple GPS redundancy that ensures well-controlled flight. This hex-copter can lift heavy payloads compared to a standard quad-copter such as high-grade colour cameras. Further, due to the nature of the nature of our inspection task, a long flight time is preferable. DJI Matrice 600 has a continuous flight

Fig. 5: DJI Matrice 600, the drone used for our flights.

Fig. 6: DJI Zenmuse X5.

TABLE III: DJI Matrice 600 Specifications.

Structure	
Drone dimensions	1668mm $\times$1518mm $\times$ 759mm
Weight	9.1 kg
Maximum take-off weights	15.1 kg
Performance	
Hovering Accuracy	Vert: $\pm$0.5 m, Horiz: $\pm$1.5 m
Maximum angular velocity	Pitch: 300°/s, Yaw: 150°/s
Maximum angle pitch	25°
Maximum speed of ascent	5m/s
Maximum speed of descent	3 m/s
Maximum wind resistance	8 m/s
Maximum altitude (above sea level)	2,500 m
Top speed (no wind)	18 m/s
Hover time	35 - 40 min (no payload)
Remote operations	
Operating frequency	5.725 GHz to 5.825 GHz 2.400 GHz to 2.483 GHz
Maximum transmission distance (unobstructed, no interference)	FCC Compliant: 5 km CE Compliant: 3.5 km
Operating Temperature	$-10°$ to 40° C

TABLE IV: Specification summary of the colour camera.

Parameter	Details
Sensor Size	17.3mm x 13mm
Sensor Type	CMOS
Effective Pixels	16×10^6
ISO Range	100 - 256,000
Max Shutter Speed	1/8000 sec
Optics	15mm F/1.7 & 45mm F/1.8
Resolution Options	4096x2160, 3840x2160 2704x1520, 1920x1080
Video format	MP4/MOV/Cinema DNG (RAW)
Storage	Micro-SD Class 10/SSD default 512GB (RAW)

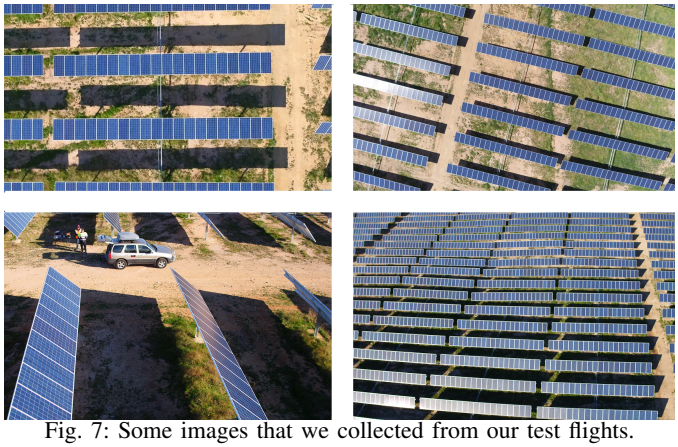


Fig. 7: Some images that we collected from our test flights.

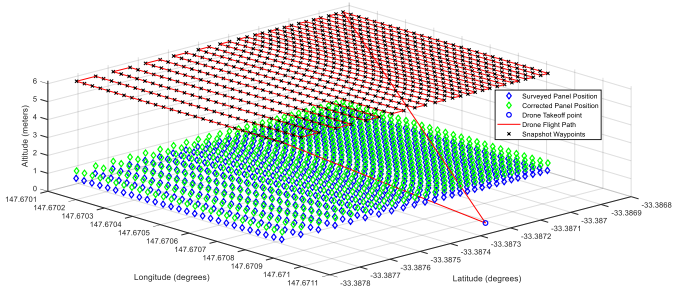


Fig. 8: Illustrative 3D drone flight path for a given solar farm.

time of 30-40 minutes and a 3.5km transmission range. The drone's specifications are summarized in Table III.

The hex-copter was equipped with a DJI Zenmuse X5, a high resolution camera. This camera, shown in Fig. 6, is capable of capturing 16 megapixel still images. The specifications of this camera are listed in Table IV.

With the assistance of our industrial partners in Australia, a large set of images were collected from several flights carried out at different times of the year. Example images are shown in Fig. 7.

During the experimental flights, the drone was autonomously controlled to follow a predefined set of waypoints and image capture points. The flight path is computed from an algorithm that uses the solar farm configuration (GPS coordinates of panels, panels height from the ground, and orientation) and includes 3D position of the drone, orientation of the drone, and the orientation of the camera gymbol. To avoid direct sun reflections, the camera view angle is set to 5 degrees with respect to the panel normal, a slight deviation from full fronto-parallel which does not substantially effect the performance of the global algorithm. An illustrative flight plan is shown in Fig. 8.

To test our algorithm, we created two PV module templates. The templates are generated by averaging multiple, hand registered, PV module images. The result averages out any module-specific appearance changes. The first template will be used for mono-modal template registration and is obtained from a greyscale version of PV module images. The second template is generated from a near-infrared (NIR) images captured using the multi-spectral Parrot Sequoia camera [23]. All images used in our experiments are obtained from calibrated sensors and are pre-processed to compensate lens distortion. The two templates are shown in Fig. 9 for the reader's reference.

In order to evaluate the accuracy of our template registration approach, we propose the intersection-of-union (IoU) criteria. This is computed by

$$\text{IoU} = \frac{|\mathcal{C}_s \cap \mathcal{C}_g|}{|\mathcal{C}_s \cup \mathcal{C}_g|} \times 100, \tag{37}$$

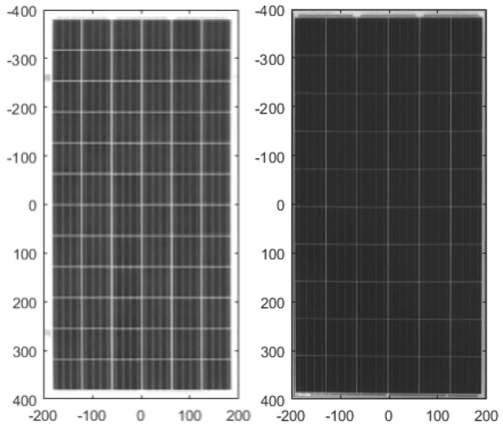


Fig. 9: The two templates used in our experiments. (Left) is the averaged template obtained from DJI Zenmuse X5, and (right) the near-infrared template captured using Parrot Sequoia.

TABLE V: Experimental Parameters.

Parameter	Meaning	Value
ℓ_{max}	Pyramid levels	3
η	Pyramidal down-sampling factor	0.5
$\mathcal{N}$	Number of histogram bins	32
δ_x^1	t_x sampling increment	2.5
δ_y^1	t_y sampling increment	2.5
δ_s^1	s sampling increment	0.02

where $\mathcal{C}_s$ and $\mathcal{C}_g$ denote the convex hulls defined from the corners of the warped template corners and the ground truth corners, respectively, and $|\cdot|$ represents the cardinality operator. The 4 corners of $\mathcal{C}_g$ are manually selected from a target PV module in $\mathcal{I}$) to evaluate the metric. The manual selection, and lack of sub-pixel precision, of the ground truth determination introduces error in the IOU estimation that may be as high as 5Also, and unless stated otherwise, the default experimental parameters used in this section are as listed in Table V.

An illustrative mono-modal registration experimental result is shown in Fig. 10. The ground truth solution selection is highlighted by the shaded area with boundary drawn in magenta. The proposed template search method gives a high IoU score (IoU = 97.82%) that lie within our estimate of human error.

A multi-modal template registration experiment was also conducted, where the Near Infra Red (NIR) template in Fig. 9 was used to register RGB images. The registration search results in a similarly high IoU score of 97.72%.

In a straightforward manner, the approach can be extended to multiple template detection, the detection of multiple PV modules in one scene. The results from this experiment are shown in Fig. 12.

V. Conclusion

This work presented a novel two-stage template registration approach based on mutual information that precisely register multi-modal sensor data of PV modules as the first step in data processing for autonomous robotic inspection of renewable infrastructure. A multi-resolution transform net is constructed to sample the global warp parameters space and searched

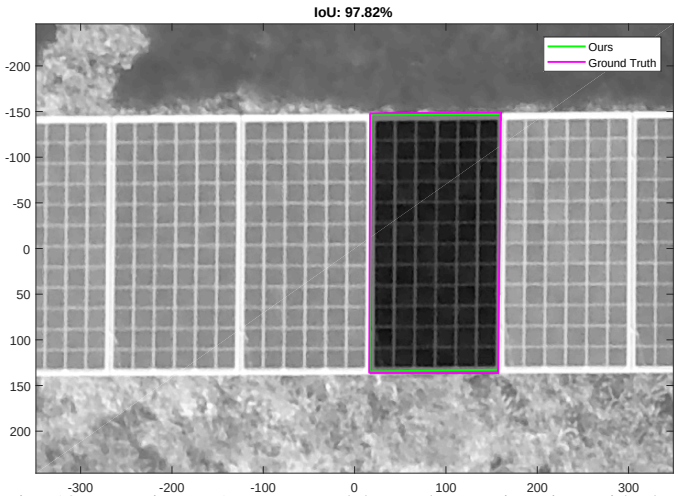


Fig. 10: Experiment 1: mono-modal template registration with the grayscale template (IoU = 97.82%).

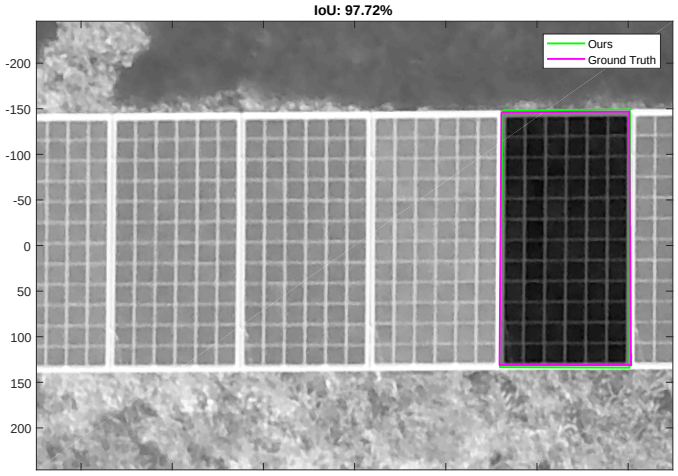


Fig. 11: Experiment 2: multi-modal template registering with the NIR template (IoU = 97.72%).

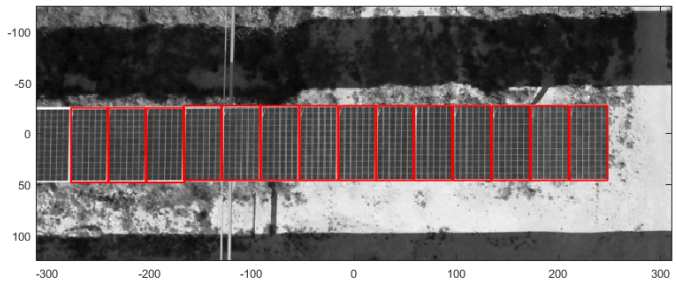


Fig. 12: Multiple PV module detection using our approach with Non-maxima suppression (only).

efficiently using a branch-and-bound algorithm. The output of the branch and bound global search was used to initialize a local trust-region descent method that determines the required image warp to subpixel accuracy. The mutual information cost criterion used is robust and can be applied to multi-modal sensor data.

Acknowledgements

This research was supported by the Australian Renewable Energy Agency (ARENA), through Grant G00853 “A robotic vision system for rapid inspection and evaluation of solar plant infrastructure”.

REFERENCES

[1] G. Spagnuolo, W. Xiao, and C. Cecati, "Monitoring, diagnosis, prognosis, and techniques for increasing the lifetime/reliability of photovoltaic systems," *IEEE Trans. Ind. Electron.*, vol. 62, pp. 7226–7227, Nov 2015.

[2] G. E. Tverberg, "Oil supply limits and the continuing financial crisis," *Energy*, vol. 37, no. 1, pp. 27 – 34, 2012. 7th Biennial Int. Workshop on Advances in Energy Studies.

[3] Ã. Netland, G. Jenssen, H. M. Schade, and A. Skavhaug, "An experiment on the effectiveness of remote, robotic inspection compared to manned," in *2013 IEEE Int. Conf. Systems, Man, and Cybernetics*, pp. 2310–2315, Oct 2013.

[4] N. Correll and A. Martinoli, "Multirobot inspection of industrial machinery," *IEEE Robot. Autom. Mag. Magazine*, vol. 16, pp. 103–112, March 2009.

[5] M. A. Abidi, R. O. Eason, and R. C. Gonzalez, "Autonomous robotic inspection and manipulation using multisensor feedback," *Computer*, vol. 24, pp. 17–31, April 1991.

[6] M. Sweatt, A. Ayoade, Q. Han, J. Steele, K. Al-Wahedi, and H. Karki, "Wifi based communication and localization of an autonomous mobile robot for refinery inspection," in *2015 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 4490–4495, May 2015.

[7] O. Araar and N. Aouf, "Visual servoing of a quadrotor uav for autonomous power lines inspection," in *22nd Mediterranean Conf. on Control and Automation*, pp. 1418–1424, June 2014.

[8] S. Dotenco, M. Dalsass, L. Winkler, T. WÃ¼rzner, C. Brabec, A. Maier, and F. Gallwitz, "Automatic detection and analysis of photovoltaic modules in aerial infrared imagery," in *2016 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, pp. 1–9, March 2016.

[9] P. B. Quater, F. Grimaccia, S. Leva, M. Mussetta, and M. Aghaei, "Light unmanned aerial vehicles (uavs) for cooperative inspection of pv plants," *IEEE J. Photovolt.*, vol. 4, pp. 1107–1113, July 2014.

[10] M. Aghaei, A. Gandelli, F. Grimaccia, S. Leva, and R. E. Zich, "Ir real-time analyses for pv system monitoring by digital image processing techniques," in *2015 Int. Conf. Event-based Control, Communication, and Signal Processing (EBCCSP)*, pp. 1–6, June 2015.

[11] S. Leva, M. Aghaei, and F. Grimaccia, "Pv power plant inspection by uas: Correlation between altitude and detection of defects on pv modules," in *2015 IEEE 15th Int. Conf. Environment and Electrical Engineering (EEEIC)*, pp. 1921–1926, June 2015.

[12] M. Aghaei, F. Grimaccia, C. A. Gonano, and S. Leva, "Innovative automated control system for pv fields inspection and remote control," *IEEE Trans. Ind. Electron.*, vol. 62, pp. 7287–7296, Nov 2015.

[13] S. Korman, D. Reichman, G. Tsur, and S. Avidan, "Fast-match: Fast affine template matching," *Int. Journal of Computer Vision*, vol. 121, pp. 111–125, Jan 2017.

[14] R. P. Brent, "An algorithm with guaranteed convergence for finding a zero of a function," *The Computer Journal*, vol. 14, no. 4, pp. 422–425, 1971.

[15] N. Dowson and R. Bowden, "Mutual information for lucas-kanade tracking (milk): An inverse compositional formulation," *IEEE Trans. Pattern Anal. Mach. Intell. Machine Intelligence*, vol. 30, pp. 180–185, Jan 2008.

[16] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 9, pp. 62–66, Jan 1979.

[17] P. Thevenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, pp. 2083–2099, Dec 2000.

[18] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. i. theory," *IEEE Trans. Signal Process.*, vol. 41, pp. 821–833, Feb 1993.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[20] E. Meinhardt-Llopis, J. SÃ¡nchez PÃ©rez, and D. Kondermann, "Horn-Schunck Optical Flow with a Multi-Scale Strategy," *Image Processing On Line*, vol. 3, pp. 151–172, 2013.

[21] J. SÃ¡nchez, "The Inverse Compositional Algorithm for Parametric Registration," *Image Processing On Line*, vol. 6, pp. 212–232, 2016.

[22] A. Dame and E. Marchand, "Second-order optimization of mutual information for real-time image registration," *IEEE Trans. Image Process.*, vol. 21, pp. 4190–4203, Sept 2012.

[23] Parrot, "Sequoia." https://tinyurl.com/ydcef5sk, 2018.