Noname manuscript No.
(will be inserted by the editor)

Edge-based multi-modal registration and application for night vision devices

Camille Sutour · Jean-François Aujol · Charles-Alban Deledalle · Baudouin Denis de Senneville

Received: date / Accepted: date

Abstract Multi-modal image sequence registration is a challenging problem that consists in aligning two image sequences of the same scene acquired with a different sensor, hence containing different characteristics. We focus in this paper on the registration of optical and infra-red image sequences acquired during the flight of a helicopter. Both cameras are located at different positions and they provide complementary informations. We propose a fast registration method based on the edge information: a new criterion is defined in order to take into account both the magnitude and the orientation of the edges of the images to register. We derive a robust technique based on a gradient ascent and combined with a reliability test in order to quickly determine the optimal transformation that matches the two image sequences. We show on real multi-modal data that our method outperforms classical registration methods, thanks to the shape information provided by the contours. Besides, results on synthetic images and real experimental conditions show that the proposed algorithm manages to find the optimal transformation in few iterations, achieving a rate of about 8 frames per second.

Keywords Multi-Modal · image sequence registration · night vision · optimization

1 Introduction

1.1 Operational context

Multi-modal image registration consists in aligning several images of a same scene acquired by different sensors, from a different point of view or at a different time. It is widely used in medical applications, for example for comparing images of the brain obtained with computer tomography (CT) to positron emission tomography (PET) or magnetic resonance (MRI) images. Multi-modal registration is also studied in remote sensing applications, for example, for the association of a synthetic aperture radar (SAR) image and an optical one. It is an important preliminary step for high level analysis such as image fusion, change detection, augmented reality, etc. A survey of most registration methods can be found in [3,20].

The goal of this paper is to perform multi-modal registration between optical image sequences obtained from a night vision device and infra-red image sequences, acquired from a helicopter, in the perspective of fusing both modalities. Optical images are obtained thanks to a

C. Sutour thanks the DGA and the Aquitaine region for funding her PhD. J.-F. Aujol acknowledges the support of the Institut Universitaire de France. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02).

C. Sutour
IMB and LaBRI, Université de Bordeaux, Talence, France,
E-mail: camille.sutour@math.u-bordeaux.fr

J.-F. Aujol, Ch.-A. Deledalle, B. Denis de Senneville
IMB, CNRS, UMR 5251, Université de Bordeaux, Talence, France,
E-mail: {jaujol,cdeledal,bdenisde}@math.u-bordeaux.fr

B. Denis de Senneville
Imaging Division, UMC Utrecht

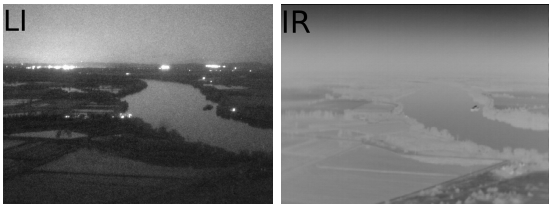

Fig. 1 Example of (left) optical (LI) and (right) infra-red (IR) images simultaneously acquired from a flying helicopter. The resolution and the information displayed are different, as well as the intensity distributions.

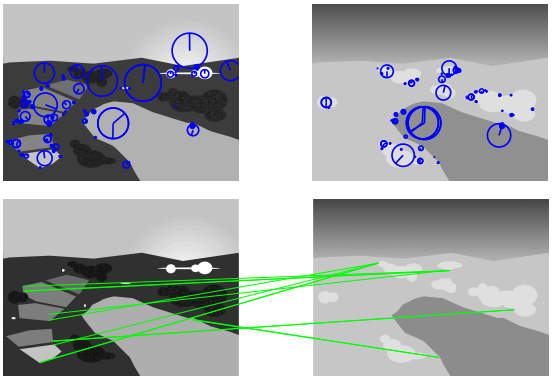

Fig. 2 Detection of SIFT features and associated descriptors on each modality, and matching of the features obtained from different modalities.

light intensifier (LI) that multiplies the photons in order to amplify the luminosity. The light intensifier is combined with a CCD camera to obtain numerical images. The LI device is located on the helmet of the pilot and the images are projected on the visor. They display natural scenes and are easily interpretable, but they suffer from classical defaults inherent to night vision devices: they are degraded by (photon count) noise and they suffer from artifacts (meshing, changes of illumination...), they are poorly contrasted while using a wild dynamic range, and they are saturated around light sources. An example is showed in Figure 1. On the other hand, infra-red (IR) images reflect the temperature of the scene. They are not easy to interpret because they do not reflect the intuitive perception of the scene, as shown on Figure 1. However, they provide precious information such as vehicles, roads and buildings because they are hot sources compared to the ground. The infra-red camera is located at the bottom of the helicopter, and can be driven by the pilot. These two video cameras observe the scene from a different angle, and they can move independently from each other, so a careful registration that takes into account both the difference of perspective and the relative movement between the two must be achieved prior to combining the information.

In the scope of this study, the optical (LI) image and the infra-red (IR) image are acquired simultaneously with the same update rate. Each time a new couple of images (LI, IR) is obtained, the goal is to register the optical image into the frame of reference of the infra-red image IR. This consists in finding the global spatial transformation T that associates each pixel of the LI image (referred to as the current image u) to its corresponding location in the IR image (referred to as the reference image v).

1.2 State of the art

Registration techniques are mostly issued from medical applications, remote sensing, or computer vision. In computer vision, registration techniques are often based on the detection and matching of special features. Among them, the scale invariant feature transform (SIFT) descriptor proposed by Lowe in [8] combines a scale invariant region detector with a descriptor based on the gradient distribution in the detected regions. These descriptors have been widely studied and adapted [2,10], combined with classical matching algorithms such as the RANSAC algorithm [5], in order to estimate the optimal transformation. The SIFT descriptors have also been adapted for example for SAR images in [4] or CT/MRI medical data [12], but the proposed SIFT implementations cannot intrinsically deal with a multi-modal framework. Indeed, even if each modality benefits from its own feature descriptors, it is not always possible to match the descriptors together due to the different information inherent to each modality [19]. Figure 2 shows on synthetic images the difficulty to associate features extracted from different sensors: even though each modality has its own features, the descriptors do not allow to perform accurate feature matching, so no transformation can be directly estimated.

An alternative to deal with different modalities consists in finding the optimal domain transformation T that maximizes a given energy. The cross-correlation metric [14,15] measures the correlation between the values of the current transformed image, denoted $u(T)$, and

the reference one v using the following formula:

$$CC(T) = \frac{\int_\Omega u_0(T(X)).v_0(X)\mathrm{d}X}{\sqrt{\int_\Omega u_0(T(X)^2\mathrm{d}X}.\sqrt{\int_\Omega v_0(X)^2\mathrm{d}X}}. \quad (1)$$

where Ω is the (continuous) image domain, and $u_0(T(X))$ and $v_0(X)$ are the centered values of the images, ie the difference between the image and the average value of the image $\overline{u(T)}$ or $\overline{v}$: $u_0(T(X)) = u(T(X)) - \overline{u(T)}$ and $v_0(X) = v(X) - \overline{v}$. However, this metric is based on the assumption that the intensities of the images to register are close up to an affine scaling, which is not the case in our multi-modal problem.

Mutual information [18,9] reflects the relation between the intensity distributions, but without any assumption regarding the nature of this relation. Mutual information is issued from information theory and the notion of entropy. It can also be interpreted in terms of a Kullback-Leibler distance:

$$I(T) = \int_{\Omega^2} p(u(T(X)), v(Y)) \cdot \log \frac{p(u(T(X)), v(Y))}{p(u(T(X))p(v(Y))} \mathrm{d}X\mathrm{d}Y, \quad (2)$$

where $p(u, v)$ is the joint probability distributions of u and v and $p(u)$ and $p(v)$ are the marginal distributions. It measures the information that one data contains about the other: the more independent u and v are, the closer the joint probability $p(u, v)$ is to the distribution $p(u) \times p(v)$. In practice, these distributions are estimated by computing the marginal and joint histograms of the values of the images to register. The registration is then performed by seeking the transformation that will maximize the mutual information between both images. However, it still requires that the intensities of both modalities are close to be in bijection. Unfortunately, this is not satisfied in our problem: some areas are highly textured on the optical image but smooth on the IR image, while some constant areas on the optical image (such as the sky or the river on Figure 1) are shaded on the IR image.

Registration can also be performed using the edge information of the images. In [17], an edge-based metric is defined in order to measure the alignment of the gradient ∇v of the reference image and the gradient $\nabla u(T)$ of the transformed version of the image to register, where T is the tested transformation. The edge alignment is evaluated at each pixel thanks to the following edge-based criterion:

$$C_S(T) = \frac{\int_\Omega w_T(X) \cos(2\Delta\theta_T(X))\ \mathrm{d}X}{\int_\Omega w_T(X)\ \mathrm{d}X} \quad (3)$$

where $w_T(X)$, $\Delta\theta_T(X)$ are based on the magnitude M and the orientation θ of the image gradients at location X:

$$\begin{aligned} w_T(X) &= M_u(T(X))M_v(X), \\ \Delta\theta_T(X) &= \theta_u(T(X)) - \theta_v(X). \end{aligned} \quad (4)$$

The $\cos(2\Delta\theta_T(X))$ in equation (3) favors the transformations that align the edge direction, regardless of the gradient orientation. Besides, when dealing with multi-modal images, some discontinuities can only appear in one of the two modalities, so the weight $w_T(X)$ favor strong edges that occur in both modalities.

In [6], a similar edge-based metric is used, based on the following quantity:

$$\omega_T(X) = \left\langle \frac{\nabla u(T(X))}{\|\nabla u(T(X))\|_\epsilon}, \frac{\nabla v(X)}{\|\nabla v(X)\|_\epsilon} \right\rangle^2 \quad (5)$$

with $\|\nabla v(X)\|_\epsilon = \sqrt{\nabla v(X)^T \nabla v(X) + \epsilon^2}$. The Normalized Gradient Fields metric is defined as follows :

$$\begin{aligned} C_H(T) &= \int_\Omega \omega_T(X)\ \mathrm{d}X \\ &= \int_\Omega \cos^2(\Delta\theta_T(X))\ \mathrm{d}X \end{aligned} \quad (6)$$

This metric uses normalized gradients; it can be expressed as a scalar product or as the cosine of the angle between the edges, regardless of the edge amplitude.

This edge-based metric can be traced back to [13] where the shape information is combined to the mutual information. In order to take into account the edges that appear in both modalities, the scalar product $\omega_T(X)$ is weighted by the minimum of the gradient magnitude:

$$G(T) = \int_\Omega \omega_T(X) \min(|\nabla u(T(X))|, |\nabla v(X)|)\ \mathrm{d}X, \quad (7)$$

then $G(T)$ is combined with mutual information in order to take into account both spatial and distribution-based information.

1.3 Contribution and organization of the paper

We propose to extend the edge-based metrics of (3) and (6) to a robust night vision framework as follows: we develop a new criterion that takes into account both the magnitude and the direction of the edges, and we maximize this criterion using a gradient ascent scheme in order to find the best transformation that will align one image with the other.

Our main contributions are the new criterion we propose, that we can express in a continuous form, and the theoretical and experimental study we conduct in order to validate the proposed model. We also develop a gradient ascent optimization combined to a temporal validation scheme that can proceed up to 8 images per second, which makes it suitable to an embedded operational registration.

The proposed model is presented in section 2, then section 3 provides an optimization scheme based on a gradient ascent, and combined with a temporal scheme that guarantees stability and robustness. Section 4 studies the performance of the metric: we show that the maximization of the proposed criterion does allow to recover the optimal transformation, both in theory and in practice, and we study the stability and robustness of this metric. We also check that the gradient ascent scheme allows to recover the optimal transformation parameters. Section 5 presents an extension of the proposed model to the general case of projective transformations and extends the gradient ascent accordingly. Finally, section 6 shows results on real data.

2 Multi-modal framework

2.1 Definition of the criterion

The current image u is registered to the reference position given by v as follows. Let M_v and θ_v be the magnitude and the orientation of the gradient of the reference image v, computed using a Sobel edge detector [16]. M_u and θ_u are defined similarly in the image to register u.

We define an edge-based metric that is adapted to the night vision framework and the characteristics of each modality, and that is easy to manipulate in an embedded operational context. We define the following criterion:

$$C(T) = \int_\Omega |\nabla u(T(X)) \cdot \nabla v(X)| \ \mathrm{d}X, \tag{8}$$

that can also be written under the following form:

$$C(T) = \int_\Omega w_T(X) \, |\cos(\Delta\theta_T(X))| \ \mathrm{d}X, \tag{9}$$

where $w_T(X)$ and $\Delta\theta_T(X)$ refer to the magnitude and orientation quantities defined in (4). This criterion favors strong edges, thanks to the amplitude ponderation, and it is insensitive to the orientation of the gradient, only to the direction, thanks to the absolute value of the cosine. It allows to take into account edges that occur in both modalities, regardless of their orientation. Besides, contrary to the metric proposed in [17], this criterion is not normalized. This makes it easier to manipulate (more stable), and it is more sensitive to the number of edges that are actually put in correspondence. Indeed, the normalized criterion performs a weighted average of the score obtained for each edge, so it measures the average edge alignment that has been performed on all the edges that occur in both modalities. On the contrary, this un-normalized criterion adds up the score of each aligned edge, so that the more edges are in correspondence the higher the criterion is. The normalized criterion might favor very precise alignments, regardless of the number of matches, while being sensitive to mismatches, whereas this un-normalized criterion might prefer slightly less precise matches, if they occur often enough.

2.2 Transformation model

In the original paper of Sun et al. [17], the criterion of (3) is optimized by performing an exhaustive search on all the possible transformation parameters, that originally consist of a translation in both directions. In the scope of our application, we have first considered for possible transformations a translation in both directions (horizontal and vertical), and a uniform zoom. We denote the zoom parameter z, and the translation parameters in the horizontal and vertical direction respectively t_1 and t_2. If $X = (x \ y \ 1)^T$ are the coordinates of the image to register (that we can also note in the

concise form $X = (x\ y)^T$), we can define the transformation matrix $T = T_{t_1,t_2,z}$ as :

$$T_{t_1,t_2,z}(X) = \begin{pmatrix} 1 & 0 & t_1 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1+z & 0 & 0 \\ 0 & 1+z & 0 \\ 0 & 0 & 1 \end{pmatrix} X = \begin{pmatrix} 1+z & 0 & t_1 \\ 0 & 1+z & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (10)$$

3 Proposed optimization scheme

3.1 Gradient ascent

Thanks to the formulation proposed in equation (8), an explicit optimization scheme is derived to maximize the proposed metric at each iteration n, by performing a gradient ascent on the transformation $T_{t_1,t_2,z}$:

$$\begin{cases} t_1^{n+1} = t_1^n + \lambda_1 \partial_{t_1} C(T_{t_1^n,t_2^n,z^n}) \\ t_2^{n+1} = t_2^n + \lambda_2 \partial_{t_2} C(T_{t_1^n,t_2^n,z^n}) \\ z^{n+1} = z^n + \lambda_3 \partial_z C(T_{t_1^n,t_2^n,z^n}) \end{cases}, \quad (11)$$

where the derivatives of the function $C(T_{t_1,t_2,z})$ are at each iteration:

$$\begin{aligned} \partial_{t_1} C(T_{(t_1,t_2,z)}) &= \int_\Omega \sigma D^2 u(T_{t_1,t_2,z}(X)) \begin{pmatrix} 1 \\ 0 \end{pmatrix} . \nabla v(X) \mathrm{d}X, \\ \partial_{t_2} C(T_{(t_1,t_2,z)}) &= \int_\Omega \sigma D^2 u(T_{t_1,t_2,z}(X)) \begin{pmatrix} 0 \\ 1 \end{pmatrix} . \nabla v(X) \mathrm{d}X, \\ \partial_z C(T_{(t_1,t_2,z)}) &= \int_\Omega \sigma D^2 u(T_{t_1,t_2,z}(X)) \begin{pmatrix} x \\ y \end{pmatrix} . \nabla v(X) \mathrm{d}X. \end{aligned} \quad (12)$$

where $\sigma = \mathrm{sign}(\nabla u(T_{t_1,t_2,z}(X)).\nabla v(X))$.

The computation of the derivatives is detailed in appendix A. The functional we seek to maximize is subject to local maxima, so the initialization is important. For the first frame, we can either perform a coarse exhaustive search as in [17], or perform several gradient ascents with different initializations and select the result that gives the best metric value. Then in practice, the sequence provides temporal regularity, so the transformation for each frame can be initialized with the parameters obtained from the previously acquired frame.

3.2 Temporal implementation

In order to accelerate the convergence, improve the stability of the algorithm and control the performance of the registration, the gradient ascent has been included into a temporal scheme

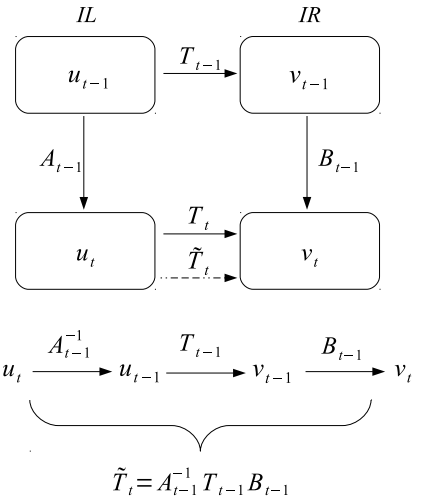

Fig. 3 Proposed temporal scheme for a fast convergence of the gradient ascent algorithm, and error control.

that uses the information from the previously registered frames to predict and control the registration for the next frames. Figure 3 displays the different steps for a fast, in-flight registration. At time $t-1$, the LI and IR images respectively called u_{t-1} and v_{t-1} are registered with transformation T_{t-1}, so that:

$$u_{t-1}(T_{t-1}X) = v_{t-1}(X). \quad (13)$$

Besides, for a single modality between u_{t-1} and u_t or v_{t-1} and v_t, movement estimation can be achieved with simpler mono-modal methods. In this case, we have chosen to use Motion 2D [11] separately on each modality, in order to estimate the transformation A_{t-1} between u_{t-1} and u_t and B_{t-1} between v_{t-1} and v_t, such that:

$$\begin{cases} u_{t-1}(A_{t-1}X) = u_t(X) \\ v_{t-1}(B_{t-1}X) = v_t(X) \end{cases} \quad (14)$$

Note that in practice in operational conditions, cameras are equipped with posture detection systems that can (roughly) estimate the movement of each camera between two acquisitions.

Thanks to these three estimations, it is possible to predict the estimated transformation $\widetilde{T}_t$ between u_t and v_t as:

$$\widetilde{T}_t = A_{t-1}^{-1} T_{t-1} B_{t-1} \quad (15)$$

This estimation $\widetilde{T}_t$ can be used as a close enough initialization at time t. This ensures that the gradient ascent will converge in few iterations, and since the initialization is reasonable it will lead to a relevant maximum.

Besides, this procedure can also be used to control the energy and prevent any divergence of the gradient ascent process. Indeed, the algorithm can be subject to local maxima and it is quite sensitive to the gradient steps for each parameter. The temporal validation can balance this sensitivity: if the energy at the end of the gradient ascent is found to be lower than the initialization, this means that the algorithm has not converged properly, so we can choose to stick to the estimation $\widetilde{T}_t$.

4 Analysis and validation of the proposed model

4.1 Theoretical analysis

The goal of this section is to study the registration from a mathematical point of view in order to show that the maximization of the proposed criterion does result theoretically in finding the optimal parameters.

We study the one dimensional case and we focus on aligning two edges when the signal to register is subject to a translation and a zoom. The registration of only one edge is subject to an aperture problem, since an edge, seen from different levels of zoom, remains the same. To remedy this, we use for a reference signal a box function $v_0(x) = 1$ if $x \in [-1, 1]$, 0 otherwise (see Figure 4).

The signal to register is then defined as $u_0(x) = v_0(ax - b)$ where a is the zoom parameter with $a > 0$ (a corresponds to the factor $1 + z$ in the transformation model described in section 2, equation (10)) and b the translation parameter.

Intuition

The reference signal described above is not differentiable in ± 1. However, its derivative ∇v_0 can be represented as the sum of two Diracs at location ± 1: $\nabla v_0(x) = \delta_{-1} - \delta_1$. We can also define $\nabla u_0(x) = a \cdot \nabla v(ax - b) = a\left(\delta_{\frac{-1+b}{a}} - \delta_{\frac{1+b}{a}}\right)$, so that the functional that we seek to maximize can be expressed in a heuristic way as:

$$\begin{aligned} F(a,b) &= \int_{\mathbb{R}} |\nabla u_0(x) \cdot \nabla v_0(x)| \, \mathrm{d}X \\ &= \int_{\mathbb{R}} a \, |\nabla v_0(ax-b) \cdot \nabla v_0(x)| \, \mathrm{d}X \\ &= \int_{\mathbb{R}} a \left| \left(\delta_{\frac{-1+b}{a}} - \delta_{\frac{1+b}{a}}\right) \cdot \left(\delta_{-1} - \delta_1\right) \right| \mathrm{d}X \\ &= \int_{\mathbb{R}} a \left(\delta_{\frac{-1+b}{a}} + \delta_{\frac{1+b}{a}}\right) \cdot \left(\delta_{-1} + \delta_1\right) \mathrm{d}X \end{aligned} \tag{16}$$

Although it is not formally correct to deal with Dirac products, one can presume how the functional is going to behave thanks to this formulation:

- Perfect match between both pairs:
  In order for both pairs of Diracs to coincide at the same time, a and b must satisfy the following conditions:

$$\begin{cases} \frac{-1+b}{a} = -1 \\ \frac{1+b}{a} = 1 \end{cases} \Leftrightarrow \begin{cases} b = 0 \\ a = 1 \end{cases} \tag{17}$$

- Match of one pair:
  For only one Dirac of ∇u_0 to coincide with one Dirac of ∇v_0, a and b need to satisfy one of the following conditions:

$$\begin{array}{rcl} \frac{1+b}{a} = 1 & \Leftrightarrow & a - b = 1 \\ \frac{-1+b}{a} = -1 & \Leftrightarrow & a + b = 1 \\ \frac{1+b}{a} = -1 & \Leftrightarrow & a + b = -1 \\ \frac{-1+b}{a} = 1 & \Leftrightarrow & b - a = 1 \end{array} \tag{18}$$

- In any other case, both pairs are separate, which leads to a null functional.

This heuristic study shows three configurations:

- One unique case ($a = 1$ et $b = 0$) for which both pairs of Diracs masses are perfectly aligned, meaning that both edges are correctly registered. Intuitively, this is when the functional is at its maximum, although it is not possible to formally evaluate its value due to the product of Diracs.
- 4 linear relations between a and b for which the signals have only one pair of edges out of two that matches. Theses relationships reflect an infinite number of local maxima, whose value is assumed to be lower than the perfect registration.
- No match between any edge, resulting in a null functional.

In order to describe mathematically the behavior of the functional and separate the global maximum from the local maxima, we study an approximation of the problem on differentiable signals that represents a differentiable approximation of the box function.

Theoretical registration

In order to deal with differentiable signals, we use the following approximation of the Heaviside function [1]:

$$H_\alpha(x) = \begin{cases} \frac{1}{2}(1 + \frac{x}{\alpha} + \frac{1}{\pi}\sin\frac{\pi x}{\alpha}) & \text{if } |x| \leq \alpha \\ 1 & \text{if } x > \alpha \\ 0 & \text{if } x < -\alpha \end{cases} \tag{19}$$

This function is differentiable, and its derivative is given by:

$$\delta_\alpha(x) = \begin{cases} \frac{1}{2\alpha}(1 + \cos\frac{\pi x}{\alpha}) & \text{if } |x| \leq \alpha \\ 0 & \text{if } |x| > \alpha \end{cases} \tag{20}$$

When $\alpha \to 0$, $\delta_\alpha \to \delta$ and $H_\alpha \to H$ where δ refers to a Dirac distribution and H the Heaviside function. In practice, α would tend to 0 to simulate a discrete edge.

Based on the previous definition of the reference signal v_0, we can now rely on its continuous approximation:

$$v(x) = H_\alpha(x+1) - H_\alpha(x-1) \tag{21}$$

which is a box function on the interval $[-1-\alpha; 1+\alpha]$. Its derivative ∇v is given by:

$$\nabla v(x) = \delta_\alpha(x+1) - \delta_\alpha(x-1) \tag{22}$$

The signal to register and its derivative become:

$$\begin{aligned} u_{a,b}(x) &= v(ax-b) \\ &= H_\alpha(ax-b+1) - H_\alpha(ax-b-1), \\ \nabla u_{a,b}(x) &= a\left(\delta_\alpha(ax-b+1) - \delta_\alpha(ax-b-1)\right). \end{aligned}$$

Note that when $a > 1$, it results in a negative zoom, which means that the size of the support is reduced by a factor a, while the height of the peaks is multiplied by a factor a, hence heightened. On the contrary when $a < 1$ it is a positive zoom: the support is stretched by a factor $1/a$ and the height of the peaks is reduced. Figure 4 shows the effect of a zoom ($a > 1$ et $a < 1$) on the function u and its derivative.

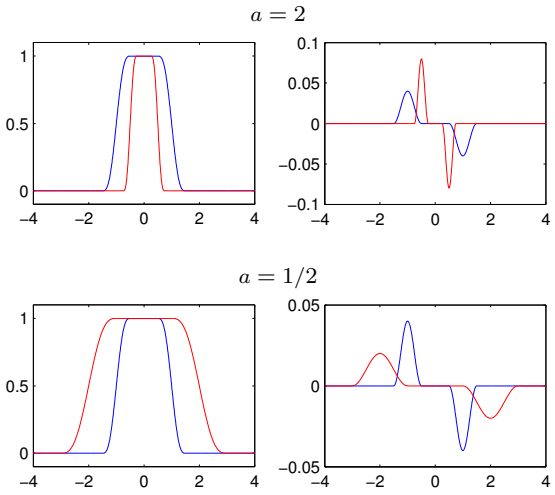

Fig. 4 Approximation with $\alpha = 0.5$ of the box function (on the left, in blue) and its derivative (on the right, in blue) and zoom with different values of a (in red).

The functional we seek to maximize becomes:

$$\begin{aligned} F(a,b) &= \int_{\mathbb{R}} |\nabla u_{a,b}(x).\nabla v(x)| \, \mathrm{d}x \\ &= a\int_{\mathbb{R}} (\delta_\alpha(x+1) + \delta_\alpha(x-1)) \\ &\quad \cdot (\delta_\alpha(ax-b+1) + \delta_\alpha(ax-b-1)) \, \mathrm{d}x \\ &= a\int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b+1) \\ &\quad +\delta_\alpha(x+1)\delta_\alpha(ax-b-1) \\ &\quad +\delta_\alpha(x-1)\delta_\alpha(ax-b+1) \\ &\quad +\delta_\alpha(x-1)\delta_\alpha(ax-b-1) \, \mathrm{d}x \\ &= F_1(a,b) + F_2(a,b) + F_3(a,b) + F_4(a,b) \end{aligned} \tag{23}$$

with:

$$\begin{aligned} F_1(a,b) &= a\int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b+1) \, \mathrm{d}x, \\ F_2(a,b) &= a\int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b-1) \, \mathrm{d}x, \\ F_3(a,b) &= a\int_{\mathbb{R}} \delta_\alpha(x-1)\delta_\alpha(ax-b+1) \, \mathrm{d}x, \\ F_4(a,b) &= a\int_{\mathbb{R}} \delta_\alpha(x-1)\delta_\alpha(ax-b-1) \, \mathrm{d}x. \end{aligned} \tag{24}$$

Each of the sub-functionals F_1,...,F_4 can be studied separately in order to determine the conditions on a and b for the integrals to be maximal, and the close form of $F(a,b)$.

Proposition 1 *The functional F can be expressed under the following form:*

$$F(a,b) = \begin{cases} \frac{3}{2\alpha} \text{ if } a = 1 \text{ and } b = 0, \\ \frac{1}{2\alpha}\left(1 + \frac{a^3}{\pi(a+1)(a-1)} \sin\frac{\pi}{a}\right) \\ \text{if } a > 1 \text{ and } \begin{cases} a+b = 1 \\ \text{or } a+b = -1 \\ \text{or } a-b = 1 \\ \text{or } a-b = -1 \end{cases} \\ \frac{1}{2\alpha}\left(a + \frac{1}{\pi(1+a)(1-a)} \sin\pi a\right) \\ \text{if } a < 1 \text{ and } \begin{cases} a+b = 1 \\ \text{or } a+b = -1 \\ \text{or } a-b = 1 \\ \text{or } a-b = -1 \end{cases} \\ 0 \text{ otherwise.} \end{cases} \tag{25}$$

Besides, it achieves its global maximum $\frac{3}{2\alpha}$ for $a = 1$ and $b = 0$.

A proof of this proposition is given in appendix B. This confirms the intuitive study conducted with the Dirac distributions in the first part, revealing a global maximum at the expected value $(a, b) = (1, 0)$ and linear subspaces of local maxima.

4.2 Assessment of the performance of the proposed criterion

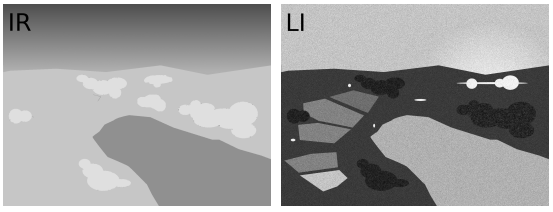

Fig. 5 Synthetic images of IR and LI modalities.

After showing that the proposed criterion is theoretically able to recover the optimal transformation parameters for the registration problem, this section aims at validating the proposed metric in practice. For the numerical evaluation, we have created synthetic images that reflect the characteristics of the involved modalities, displayed on Figure 5.

4.2.1 Study of the performances

First, we study on synthetic images the ability of the proposed criterion to find the optimal transformation parameters. We simulate some transformations with known translation and zoom parameters, then we estimate the transformation using an exhaustive search on the 3-dimensional search parameters space, that we restrict (given the images to register, and for obvious computational reasons) to a positive zoom whose coefficient z is set between -0.4 and 0.4, and translation parameters that do not exceed 40 pixels in each direction. We also study its robustness to noise, since the optical images are corrupted by a strong non-Gaussian noise: we add Poisson noise to the LI images in order to reach a PSNR of about 18dB, to reflect the natural degradations of this modality in night vision conditions. Then we compare the performances of the edge-based metrics to the classical methods described in section 1.2: the cross-correlation metric (CC) defined in equation (1), the mutual information (MI) computed from equation (2) and the combined edge-based/mutual information (MI-G) based on equation (7).

This evaluates the ability of each metric to attain its global maximum with the optimal transformation. Tables 1 and 2 display the estimated parameters with each metric on several simulated transformations, on clear then Poisson-corrupted images. The results show that the edge-based methods provide a more reliable estimation, that is also more robust to noise.

4.2.2 Study of the stability

Since we seek to perform an optimization scheme for the search for the optimal parameters, the stability of the metric and its sensitivity to local maxima is a crucial point.

The next experiment consists in evaluating the stability of the different metrics regarding the variation of one parameter. We simulate a known transformation on the synthetic images, and we compute the criterion while testing only one parameter, the other two being fixed to the correct value. Figures 6 and 7 display the evolution of the metrics as a function of the horizontal translation parameter t_1, while the vertical translation parameter t_2 and the zoom parameter z are fixed, in the case of the cross-correlation metric, the mutual information metric, the edge-based metric of Sun et al. [17], the combined edge-based/mutual information [13], the normalized Gradient Fields [6] and the proposed metric. On the top, the experiment was

Noiseless images				
Parameters	CC	MI	MI-G	Proposed metric
$[0, 0, 0]$	$\surd$	$[-16, 0, 0.05]$	$\surd$	$\surd$
$[16, -8, 0.1]$	$\surd$	$[24, 6, 0.05]$	$\surd$	$\surd$
$[32, 0, 0.15]$	$[32, -4, 0.10]$	$[26, -2, 0.05]$	$\surd$	$\surd$
$[24, -32, 0.3]$	$\surd$	$\surd$	$\surd$	$\surd$
$[8, -8, 0.4]$	$\surd$	$\surd$	$\surd$	$\surd$
$[-40, 40, 0.4]$	$[-38, 36, 0.35]$	$\surd$	$\surd$	$\surd$

Table 1 Estimated transform parameters $[t_1, t_2, z]$ obtained on synthetic images with the cross-correlation (CC), the mutual information (MI), the combined edge-based/mutual information metric (MI-G) and the proposed metric. $\surd$ refers to a correct estimation.

Noisy LI image, Poisson noise				
Parameters	CC	MI	MI-G	Proposed metric
$[0, 0, 0]$	$[2, -2, 0.05]$	$[2, 4, -0.05]$	$[-4, -14, -0.05]$	$\surd$
$[16, -8, 0.1]$	$[16, -12, 0.05]$	$\surd$	$\surd$	$\surd$
$[32, 0, 0.15]$	$[32, -4, 0.10]$	$[32, 2, 0.15]$	$[32, 2, 0.15]$	$\surd$
$[24, -32, 0.3]$	$\surd$	$\surd$	$\surd$	$\surd$
$[8, -8, 0.4]$	$[10, -6, 0.4]$	$\surd$	$\surd$	$\surd$
$[-40, 40, 0.4]$	$[-36, 38, 0.35]$	$\surd$	$\surd$	$\surd$

Table 2 Estimated transform parameters $[t_1, t_2, z]$ obtained on synthetic images with the cross-correlation (CC), the mutual information (MI), the combined edge-based/mutual information metric (MI-G) and the proposed metric. The LI image has been corrupted by a Poisson noise, so that its initial PSNR is around 18dB. $\surd$ refers to a correct estimation.

conducted on noiseless images with optimal parameters $[t_1, t_2, z] = [0, 0, 0]$, while on the bottom the LI image was corrupted by Poisson noise (initial PSNR $\approx$ 20dB), and with true parameters $[t_1, t_2, z] = [24, -32, 0.3]$.

These figures illustrate that the non-normalized edge-based metrics (bottom line) are more stable, hence more prone to optimization. Both studies on performance and stability lead us to confirm the theoretical results and to validate the proposed metric.

4.3 Study of the optimization scheme

The above theoretical and experimental studies have sought to validate the proposed metric in terms of performance, robustness and relevance. We have shown that the maximization of the proposed metric does lead to the optimal transformation, both in theory and in practice. We have also studied its robustness to noise and its behavior regarding the evolution of one transformation parameter, and we have demonstrated its ability to find the optimal transformation when searching the whole parameter space.

Now that the performance of the functional has been validated, we focus on the optimization scheme that we have developed. Indeed, even though the functional is not concave, we can study its ability to attain the global maximum within a gradient ascent scheme, provided that the initialization is close enough to the solution. In our night vision context, we benefit from a video flux that ensures that the registration performed for the previous frames is a good guess for the next couple of images.

Figure 8 displays the map of the metric computed for a fixed zoom parameter. The black line shows the path of the estimated translation parameters t_1 and t_2 at each iteration of the gradient ascent scheme. We can see that each step does maximize the metric, and that the algorithm converges towards the optimum. Besides, even though the criterion is not strictly concave, this map shows that around the optimum the metric behaves well, which guarantees that with a "good" initialization the gradient ascent will converge to the global maximum.

Such a "good" initialization is actually guaranteed by the fact that we treat videos, so we can benefit from the estimation of the frame before as described in section 3.2. Based on the synthetic images, we have simulated movement on the IR image with the translation parameters t_1 and t_2 between -40 and +40 pixels, and the zoom parameter z between 0 and 0.4, evolving with time. Then the registration with the LI image was performed using the gradient ascent for each frame. We start the first frame with a coarse exhaustive search, and we use the estimated parameters for each frame as an initialization for the next one. This guarantees that the initialization is not too far from the opti-

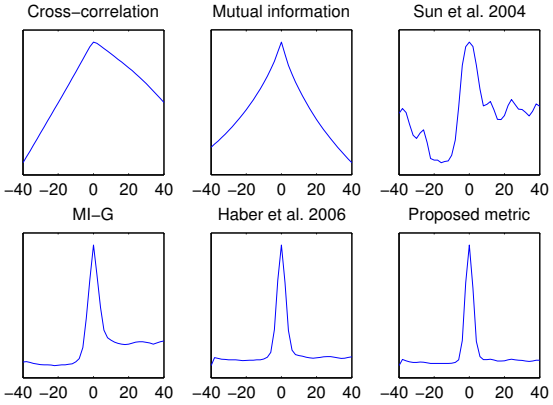

Fig. 6 Evolution of the metrics as a function of the horizontal translation parameter t_1, the other two being fixed to the optimal value. From left to right, top to bottom: Cross-correlation (1), Mutual information (2), normalized edge-based metric [17], combined Edge-based/Mutual information [13], Normalized Gradient Fields [6] and proposed metric (8). Optimal parameters: $[t_1, t_2, z] = [0, 0, 0]$.

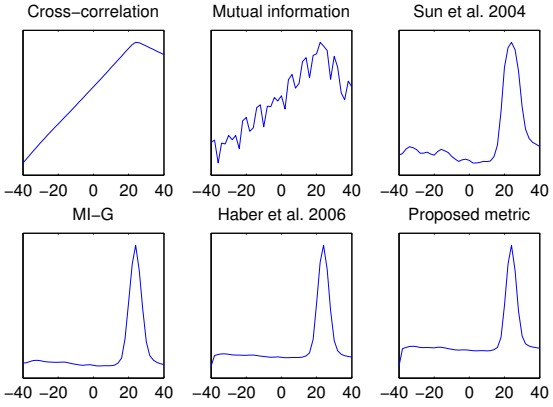

Fig. 7 Evolution of the metrics as a function the horizontal translation parameter t_1, the other two being fixed to the optimal value. The LI image was corrupted by Poisson noise in order to attain a PSNR of about 20dB. Optimal parameters: $[t_1, t_2, z] = [24, -32, 0.3]$.

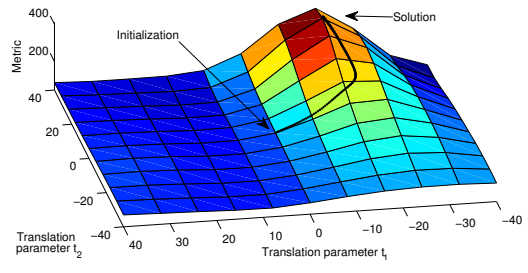

Fig. 8 Similarity map for a range of translation parameters, at a fixed zoom, and estimated parameters at each iteration of the gradient ascent. We can see that the optimization scheme does maximize the metric, and that it converges towards the optimum.

mum, and it allows to converge in a limited number of iterations. Figure 9 shows the evolution of each simulated parameter with time (blue line), and the red stars show the estimation of these parameters for each frame. The algorithm performs well on all the sequence. In fact, the average error on the estimation is less than 1 pixel for the translation and 0.003 for the zoom. These experiments show that maximizing the proposed metric is relevant since it does lead to the optimum transformation, and that the proposed gradient ascent scheme is successful.

5 Projective model

When the helicopter flies at high altitude, the assumption that the transformation between the two modalities can be modeled by a uniform zoom and a translation (or more generally by an affine transformation) can be verified. However, when the helicopter flies at lower altitude, the perspective is different between both cameras, so a projective model has to be adopted.

5.1 Projective geometry

A projective transformation [7] is described by the homography matrix

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \tag{26}$$

that has 8 degrees of freedom.

The equation of the transformation is:

$$\begin{pmatrix} wx' \\ wy' \\ w \end{pmatrix} = H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11}x + h_{12}y + h_{13} \\ h_{21}x + h_{22}y + h_{23} \\ h_{31}x + h_{32}y + 1 \end{pmatrix}, \tag{27}$$

then we revert to x' and y' by normalizing by w :

$$\begin{cases} x' = \frac{h_{11}x+h_{12}y+h_{13}}{h_{31}x+h_{32}y+1} \\ y' = \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+h_{32}y+1}. \end{cases} \tag{28}$$

We can simplify the expression using only the first two coordinates:

$$X' = HX = \begin{pmatrix} \frac{h_{11}x+h_{12}y+h_{13}}{h_{31}x+h_{32}y+1} \\ \\ \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+h_{32}y+1} \end{pmatrix} \tag{29}$$

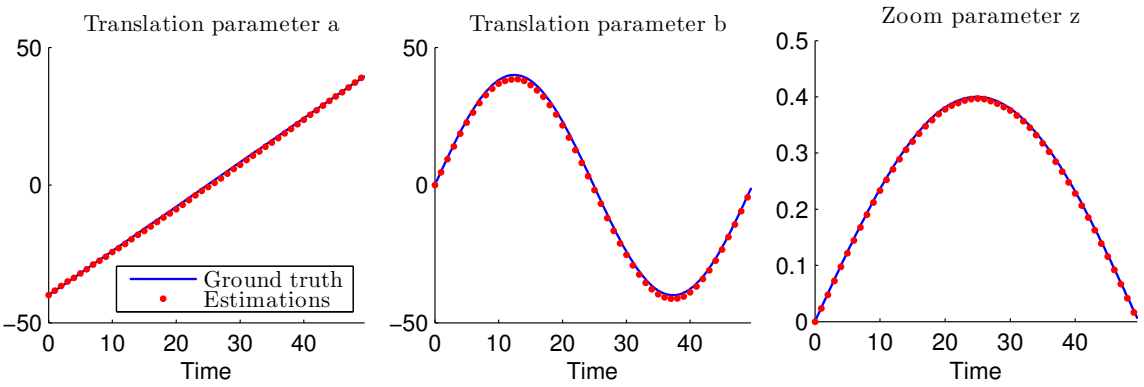

Fig. 9 Evolution of each parameter (t_1, t_2, z) of the transformation during the sequence and estimation (in red stars) computed using the gradient ascent. Some of the errors are due to the fact the evolution of the parameters is continuous, so the values are not integers, which generates approximations in the transformation.

This type of transformation is a generalization of the affine model, and it includes the transformation model we considered until then, but also the rotations and the changes of perspective.

Even though it is possible to restrain the space of the sought parameters, an exhaustive search would still require a higher number of dimensions which makes it computationally difficult in real-time. The gradient ascent then takes all its meaning in the extended problem.

5.2 Gradient ascent

The functional $F = F(H)$ that we seek to maximize now relies on 8 parameters, so we have to compute 8 partial gradients regarding each parameter. We note H_{11} the variation on parameter h_{11} :

$$H_{11} = \begin{pmatrix} h_{11}+\alpha & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \tag{30}$$

so

$$H_{11}X = \begin{pmatrix} \frac{(h_{11}+\alpha)x+h_{12}y+h_{13}}{h_{31}x+h_{32}y+h_{33}} \\ \\ \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+h_{32}y+h_{33}} \end{pmatrix}$$
$$= HX + \alpha \begin{pmatrix} \frac{x}{h_{31}x+h_{32}y+h_{33}} \\ \\ 0 \end{pmatrix} \tag{31}$$

We can write in a similar way the variations $H_{12}, ..., H_{23}$.

For the parameters that intervene in the denominator, we need to perform a linearization and we have:

$$H_{31}X = \begin{pmatrix} \frac{h_{11}x+h_{12}y+h_{13}}{(h_{31}+\alpha)x+h_{32}y+h_{33}} \\ \\ \frac{h_{21}x+h_{22}y+h_{23}}{(h_{31}+\alpha)x+h_{32}y+h_{33}} \end{pmatrix}, \tag{32}$$

which leads to the following result:

$$H_{31}X = HX - \alpha \frac{x}{h_{31}x + h_{32}y + h_{33}} HX + o(\alpha)$$
$$= HX \left(1 - \alpha \frac{x}{h_{31}x + h_{32}y + h_{33}}\right) + o(\alpha)$$

And also:

$$H_{32}X = \begin{pmatrix} \frac{h_{11}x+h_{12}y+h_{13}}{h_{31}x+(h_{32}+\alpha)y+h_{33}} \\ \\ \frac{h_{21}x+h_{22}y+h_{23}}{h_{31}x+(h_{32}+\alpha)y+h_{33}} \end{pmatrix}$$
$$= HX - \alpha \frac{y}{h_{31}x + h_{32}y + h_{33}} HX + o(\alpha).$$

We can then re-inject each of this variation calculation in the computation of the metric, leading to a 8-dimensional gradient.

6 Results

6.1 Computational time and implementation

The proposed method has been implemented first on Matlab, then in C^{++} and GPU to accelerate the registration time. The initial LI images are of size 1600 × 1200 pixels and the IR images are of size 768 × 576. First we resize the images to the same dimensions (using bicubical interpolation), then we perform a downsampling in order to reduce the size of the images, hence the computational time.

For an exhaustive search resolution, we need to compute the value of the metric for each

	Exhaustive search			Gradient ascent (Zoom/Translation)			Gradient ascent (Projective)		
	Matlab	C^{++}	GPU	Matlab	C^{++}	GPU	Matlab	C^{++}	GPU
Iteration time (ms)									
1600 × 1200	756.37	153.15	7.46	1455.71	195.01	9.38	1657.97	255.23	12.59
800 × 600	175.53	37.26	2.99	328.52	44.56	3.28	394.28	56.36	4.91
400 × 300	46.5	9.19	0.998	88.36	11.38	1.22	99.05	15.09	1.66
Total registration time (s)									
	≈ 10000 iterations			≈ 100 iterations			≈ 100 iterations		
1600 × 1200	7563.7	1531.5	74.6	145.571	19.501	0.938	165.797	25.523	1.259
800 × 600	1755.3	372.6	29.9	32.852	4.456	0.328	39.428	5.636	0.491
400 × 300	465	91.9	9.98	8.836	1.138	0.122	9.905	1.509	0.166
Frame rate (Hz)									
1600 × 1200	0.0001	0.0007	0.0134	0.0069	0.0513	**1.0661**	0.0060	0.0392	**0.7943**
800 × 600	0.0006	0.0027	0.0334	0.0304	0.2244	**3.0488**	0.0254	0.1774	**2.0367**
400 × 300	0.0022	0.0109	0.1002	0.1132	0.8787	**8.1967**	0.1010	0.6627	**6.0241**

Table 3 Registration time computed for the exhaustive search, the standard gradient ascent scheme (dealing with only a zoom and a translation), and the projective gradient ascent scheme, depending on the size of the image and the implementation. A GPU implementation of the gradient ascent scheme allows to perform registration in less than a second.

tested set of parameters, which implies applying the associated transformation to the current image, then computing the metric (which involves calculating the gradient of each image). This step is repeated for every set of parameters of the search space, that includes at least 10000 possibilities (in the non-projective case)!

For a gradient ascent resolution, for each iteration step the registered image is computed in order to evaluate the gradient of the metric, which also involves computing the image gradients. Experiments have shown that 100 iterations allow the gradient ascent to converge, and this number can even be reduced when associated to a temporal scheme as in section 3.2 where the initialization is refined by a monomodal registration.

The gradient ascent scheme is all the more interesting when the number of parameters to optimize becomes important, for example in the projective case. In order to illustrate the computational complexity involved with each resolution method, we display in table 3 the computational time as a function of the image size (that depends on the down-sampling factor) needed for one step of the resolution: either one iteration of the gradient ascent ascent or one computation of the metric for one set of parameters. Then by taking into account the average number of iterations needed (number of iteration steps for the gradient ascent scheme or size of the parameters search space), it gives an indication of the registration time for one image, depending on the size. The computational time can then expressed in terms of frame rate, ie the number of frames that can be processed in a second, that is shown to be up to 8 frames per second.

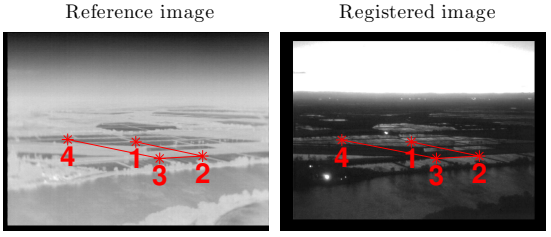

Fig. 10 Example of control points selected on the reference IR image, and associated points localized on the registered image.

Img	Orig.	CC	MI	MI-G	Edge
1	31.48	42.34	9.89	9.50	**8.44**
2	18.38	20.75	6.02	4.22	**3.30**
3	24.91	26.43	7.67	5.13	**4.97**

Table 4 Average distance between the pixel coordinates of the control points from the reference image to the image to register and to the registered image, using either the cross-correlation metric, the mutual information, the combined edge-based/mutual information or the proposed metric.

6.2 Experimental validation on real experimental conditions

In order to evaluate the different metrics on real images without knowing the optimal transformation, we have developed a method based on manually selected points. We select on the reference IR image and the LI image to register four pairs of characteristic points (the whole difficulty being to find reference points that can

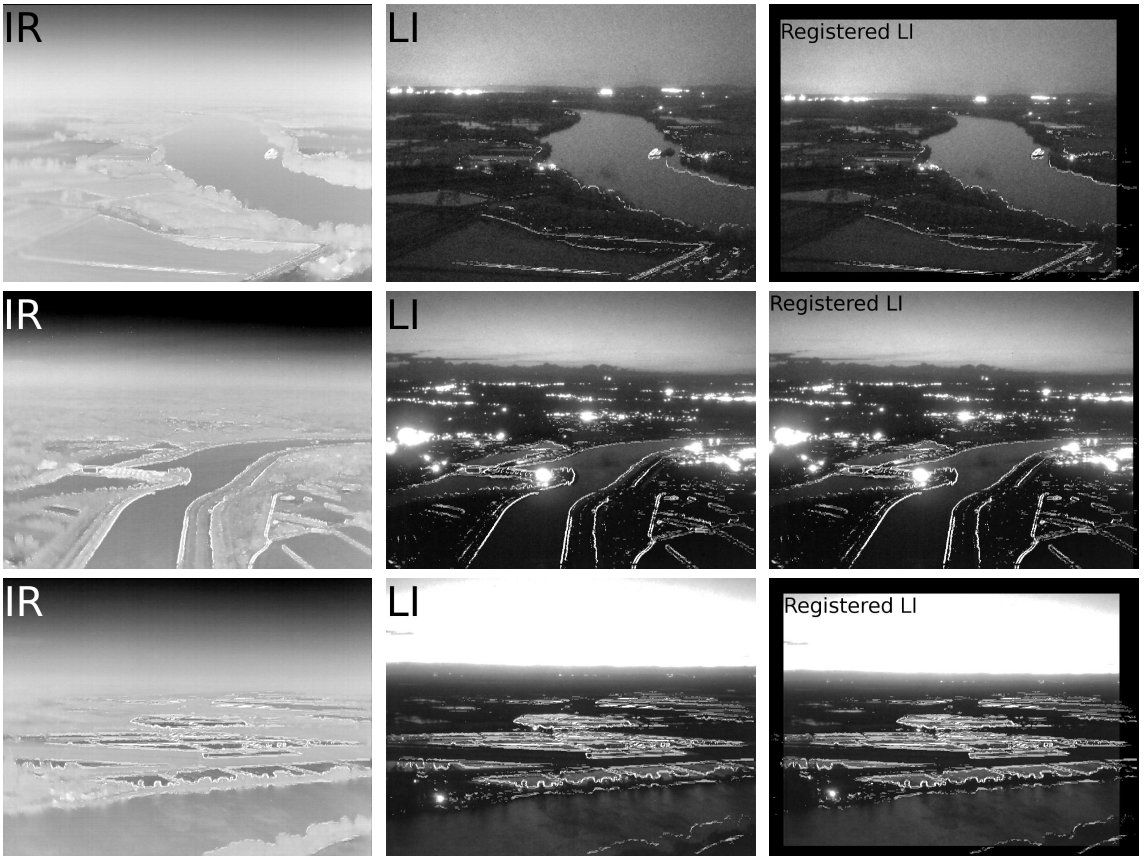

Fig. 11 Results obtained on a real data sets using the proposed metric. Edges extracted from the IR images are superimposed on each image for visual assessment of the quality of the registration. The complete registered sequences are available at http://image.math.u-bordeaux1.fr/Registration.

be identified in both images), and we measure the distance (in pixels) between the pixel coordinates for each pair of points. Then we measure the distance between the pixel coordinates from the reference image and those from the registered image. We repeat this measurement for a short sequence of images, then we average the distances to produce an average registration error (in pixels). Figure 10 displays an example of characteristics points manually selected on the reference IR image, and the corresponding points in the registered image. This experiment has been conducted on images registered with the cross-correlation, the mutual information and our proposed method. Table 4 displays the average distance between the pixel coordinates of the reference image and the registered ones using different registration metrics (cross-correlation, mutual information, combined edge-based/mutual information and proposed metric). The edge-based method is shown to perform a more accurate registration, the remaining errors being also due to the difficulty to select control points that are in perfect correspondence between the two modalities.

6.3 Registration on real data

Figure 11 displays an infra-red image, an optical image and the registered optical image in the IR coordinates issued from three different sequences. For each sequence, the images have been resized to the common resolution of 1024×768 pixels, then down-sampled by a factor 2. The parameters for the gradient ascent have been initialized using a coarse exhaustive search, then the gradient steps are set to $1e-4$ for the translation parameters and $1e-8$ for the zoom parameters. These steps have been manually optimized, but they are fixed for all the data we have tested. In order to illustrate the accuracy of the registration, we have performed an edge detection on the IR image, and we have printed these edges on the optical images, to check that the edges are correctly aligned. The

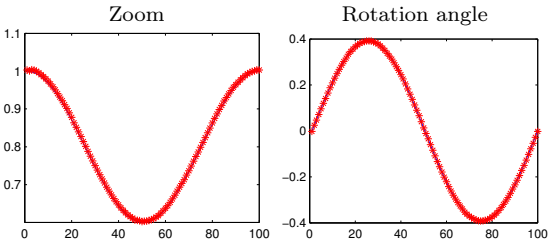

Fig. 12 Simulation of a rotating LI image sequence with varying zoom and rotation angle parameters (blue line) and estimation for the registration (red stars).

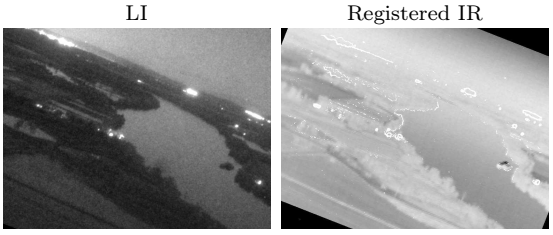

Fig. 13 Example of a transformed LI image subjected to a zoom and a rotation, and the associated registered IR image. The edges of the LI image have superimposed onto the IR one in order to illustrate the accuracy of the registration.

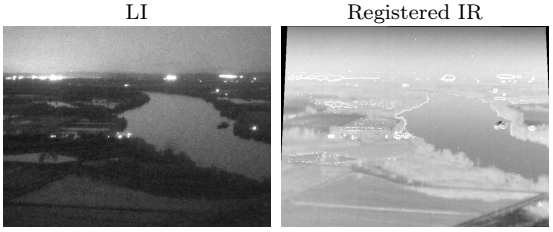

Fig. 14 Example of a transformed LI image subjected to a zoom, a translation in both directions and a horizontal projection, and the associated registered IR image. The edges of the LI image have superimposed onto the IR one in order to illustrate the accuracy of the registration.

complete registered sequences are available for download[1].

6.4 Projective simulations

We have simulated transformations on already registered (real) images. Since the projective geometry includes rotations and changes of perspective, we have simulated such movements on a LI image sequence, then performed the registration of the IR image for each frame. The knowledge of the true transformation parameters allows to check that the estimation is accurate.

[1] http://image.math.u-bordeaux1.fr/Registration

Figure 12 displays the simulated and estimated zoom and rotation angle parameters during the sequence. Figures 13 and 14 show the transformed LI image subjected to either a zoom and a rotation or a zoom, translation and horizontal projection, and the corresponding IR registered images.

These figures illustrate that the general projective model can accurately encompass all kinds of transformations that are likely to be encountered in-flight.

7 Conclusion and perspectives

In this paper, we have presented a new multisensor registration method based on the edge alignment principle. We have developed a new algorithm that aligns the edges that appear in both modalities by performing a gradient ascent scheme that provides a fast resolution. Coupled with a temporal implementation that ensures stability and provides error control, our proposed method is shown to be robust and fast compared to a standard resolution with an exhaustive search, and the algorithm can proceed up to 8 frames per second. Theoretical and experimental studies show that the criterion is relevant and liable, and results on real data validate the night vision application.

This model is well adapted to the night vision operational context, and its general properties make it suitable for any multi-modal application.

A Computation of the gradient of Functional (12)

We focus on the continuous form of the functional:
$F(T) = \int_\Omega |\nabla u(T(X)).\nabla v(X)|\, \mathrm{d}X$,
where T is the transformation we seek to optimize.

If we define a small displacement S, we have:
$F(T+S) = \int_\Omega |\nabla u(T(X)+S(X)).\nabla v(X)|\, \mathrm{d}X$
and:
$\nabla u(TX+SX)) = \nabla u(TX) + D^2u(TX)(SX) + o(S)$.
We have:

$$F(T+S) = \int_\Omega |\nabla u(T(X)+S(X)).\nabla v(X)|\, \mathrm{d}X \tag{33}$$

Using the first order expansion, we have:

$$\begin{aligned}&\int_\Omega |\nabla u(T(X)).\nabla v(X)\\&+D^2u(TX)(SX).\nabla v(X)\big|\,\mathrm{d}X\\&=\int_\Omega |\nabla u(T(X)).\nabla v(X)|\\&\times\left|1+\frac{D^2u(TX)(SX).\nabla v(X)}{\nabla u(T(X)).\nabla v(X)}\right|\mathrm{d}X\\&=\int_\Omega |\nabla u(T(X)).\nabla v(X)|\\&\times\left(1+\frac{D^2u(TX)(SX).\nabla v(X)}{\nabla u(T(X)).\nabla v(X)}\right)\mathrm{d}X\end{aligned}\tag{34}$$

so that

$$\begin{aligned}&F(T+S)=F(T)\\&+\int_\Omega \sigma D^2u(TX)(SX).\nabla v(X)\mathrm{d}X+o(S).\end{aligned}\tag{35}$$

with $\sigma=\mathrm{sign}(\nabla u(T(X)).\nabla v(X))$. We now focus on a variation on each parameter t_1,t_2,z and we see the functional F as a function of each parameter:
$F(t_1+\alpha,t_2,z)-F(t_1,t_2,z)=\alpha\partial_1F(t_1,t_2,z)$
and we denote by T_α the perturbation on T of α on the first parameter t_1, ie:

$T_\alpha=\begin{pmatrix}1+z & 0 & (t_1+\alpha)\\ 0 & 1+z & t_2\\ 0 & 0 & 1\end{pmatrix}$, and we have:

$$T_\alpha X=TX+\alpha\begin{pmatrix}1\\0\\0\end{pmatrix}$$

Hence,

$\nabla u(T_\alpha X))=\nabla u(TX)+\alpha D^2u(TX)\begin{pmatrix}1\\0\end{pmatrix}+o(\alpha)$ and:

$$\begin{aligned}&F(t_1+\alpha,t_2,z)-F(t_1,t_2,z)\\&=\alpha\int_\Omega \sigma D^2u(TX)\begin{pmatrix}1\\0\\0\end{pmatrix}.\nabla v(X)+o(\alpha)\end{aligned}\tag{36}$$

and we deduce that:

$$\partial_1F(t_1,t_2,z)=\int_\Omega \sigma D^2u(TX)\begin{pmatrix}1\\0\end{pmatrix}.\nabla v(X)\tag{37}$$

We obtain an analogous result for the second parameter t_2:

$$\partial_2F(t_1,t_2,z)=\int_\Omega \sigma D^2u(TX)\begin{pmatrix}0\\1\\0\end{pmatrix}.\nabla v(X)\tag{38}$$

For the zoom parameter z, we consider:

$T_\gamma=\begin{pmatrix}1+z+\gamma & 0 & t_1\\ 0 & 1+z+\gamma & t_2\\ 0 & 0 & 1\end{pmatrix}$, so:

$$T_\gamma X=TX+\gamma\begin{pmatrix}x\\y\\0\end{pmatrix}$$

Hence,

$\nabla u(T_\gamma X))=\nabla u(TX)+\gamma D^2u(TX)\begin{pmatrix}x\\y\\0\end{pmatrix}+o(\gamma)$
and:

$$\begin{aligned}&F(t_1,t_2,z+\gamma)-F(t_1,t_2,z)\\&=\gamma\int_\Omega \sigma D^2u(TX)\begin{pmatrix}x\\y\\0\end{pmatrix}.\nabla v(X)+o(\gamma)\end{aligned}\tag{39}$$

So we have:

$$\partial_3F(t_1,t_2,z)=\int_\Omega \sigma D^2u(TX)\begin{pmatrix}x\\y\\0\end{pmatrix}.\nabla v(X)\tag{40}$$

Putting those three differentials together, we obtain an explicit form for:

$$\nabla F(t_1,t_2,z)=\begin{pmatrix}\partial_1F(t_1,t_2,z)\\ \partial_2F(t_1,t_2,z)\\ \partial_3F(t_1,t_2,z)\end{pmatrix}.\tag{41}$$

B Proof of proposition 1: Theoretical analysis of the criterion

We have expressed the functional $F(a,b)$ that we seek to maximize as the sum of four sub-functionals:

$$\begin{aligned}F_1(a,b)&=a\int_{\mathbb{R}}\delta_\alpha(x+1)\delta_\alpha(ax-b+1)\,\mathrm{d}x,\\F_2(a,b)&=a\int_{\mathbb{R}}\delta_\alpha(x+1)\delta_\alpha(ax-b-1)\,\mathrm{d}x,\\F_3(a,b)&=a\int_{\mathbb{R}}\delta_\alpha(x-1)\delta_\alpha(ax-b+1)\,\mathrm{d}x,\\F_4(a,b)&=a\int_{\mathbb{R}}\delta_\alpha(x-1)\delta_\alpha(ax-b-1)\,\mathrm{d}x.\end{aligned}\tag{42}$$

We can study each sub-functional separately in order to determine the conditions on a and b for each of them to be maximal.

Note that the parameter α that represents the width of the peaks is meant to tend to 0. The peaks issued from the derivative of the reference signal v are located at $+1$ and -1, and

their support is $[\pm 1-\alpha, \pm 1+\alpha]$. For the transformed signal $u_{a,b}$, the peaks are located in $\frac{\pm 1+b}{a}$ and the support is $[\frac{\pm 1+b-\alpha}{a}, \frac{\pm 1+b+\alpha}{a}]$, of half-width α/a.

Hence, when α tends to 0, the width of each peak tends to 0 (for the transformed signal, this implies that $a > 0$, which is relevant in practice).

This remark simplifies the problem: we can consider that if the centers of the peaks are not perfectly aligned, then it is possible to consider a small enough α such that the supports are disjoints. Hence, we split the proof into five steps and the study of the 4 sub-functionals is limited to the conditions on a and b for the centers of the peaks to be aligned.

Step 1:

$F_1(a,b) = a \int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b+1)\ \mathrm{d}x$

The support of $\delta_\alpha(x+1)$ is $[-1-\alpha; -1+\alpha]$, centered in -1, and the support of $\delta_\alpha(ax-b+1)$ is $[\frac{-1+b-\alpha}{a}; \frac{-1+b+\alpha}{a}]$, centered in $-1+b/a$.

For the function F_1 to be non-null, we solve:

$$\frac{-1+b}{a} = -1 \Leftrightarrow a+b=1 \tag{43}$$

Besides, when condition (43): $a+b=1$ is fulfilled, F_1 can be expressed in closed form:

- $a > 1$:
  If $a > 1$, the half-size of the support of $u_{a,b}$ is $\alpha/a < \alpha$, so the computation of F_1 is restricted to the interval $[-\alpha/a, \alpha/a]$, and we have:

$$\begin{aligned} F_1(a,b) &= a\int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b+1)\ \mathrm{d}x, \\ &= \frac{a}{4\alpha^2}\int_{-\frac{\alpha}{a}}^{\frac{\alpha}{a}} \left(1+\cos\frac{\pi x}{\alpha}\right)\cdot\left(1+\cos\frac{\pi a x}{\alpha}\right)\mathrm{d}x, \\ &= \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right). \end{aligned} \tag{44}$$

- $a < 1$:
  If $a < 1$, the half-size of the support of $u_{a,b}$ is $\alpha/a > \alpha$, so the computation of F_1 is restricted to the interval $[-\alpha, \alpha]$, and we have:

$$\begin{aligned} F_1(a,b) &= a\int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b+1)\ \mathrm{d}x, \\ &= \frac{a}{4\alpha^2}\int_{-\alpha}^{\alpha} \left(1+\cos\frac{\pi x}{\alpha}\right)\cdot\left(1+\cos\frac{\pi a x}{\alpha}\right)\mathrm{d}x, \\ &= \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right). \end{aligned} \tag{45}$$

- $a = 1$:
  If $a = 1$, the condition (43): $a+b=1$ implies that $b = 0$, so $u = v$, and we have:

$$F_1(1,0) = \int_{-\alpha}^{\alpha}\left[\frac{1}{2\alpha}\left(1+\cos\frac{\pi x}{\alpha}\right)\right]^2 \mathrm{d}x = \frac{3}{4\alpha} \tag{46}$$

Conclusion :

$$F_1(a,b) = \begin{cases} \frac{3}{4\alpha} \\ \text{if } a=1 \text{ and } b=0 \\ \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right) \\ \text{if } a+b=1 \text{ and } a>1 \\ \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right) \\ \text{if } a+b=1 \text{ and } a<1 \\ 0 \text{ otherwise.} \end{cases} \tag{47}$$

Step 2:

$F_2(a,b) = a \int_{\mathbb{R}} \delta_\alpha(x+1)\delta_\alpha(ax-b-1)\ \mathrm{d}x$

An analogous study on the support of $\delta_\alpha(x+1)$ and $\delta_\alpha(ax-b-1)$ leads to solving the following conditions on a and b for the support to intersect:

$$\frac{1+b}{a} = -1 \Leftrightarrow a+b=-1 \tag{48}$$

When condition (48): $a+b=-1$ is satisfied, F_2 can be computed in a similar way to F_1:

$$F_2(a,b) = \begin{cases} \frac{3}{4\alpha} \\ \text{if } a=1 \text{ and } b=-2 \\ \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right) \\ \text{if } a+b=-1 \text{ and } a>1 \\ \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right) \\ \text{if } a+b=-1 \text{ and } a<1 \\ 0 \text{ otherwise.} \end{cases} \tag{49}$$

Step 3:

$F_3(a,b) = a \int_{\mathbb{R}} \delta_\alpha(x-1)\delta_\alpha(ax-b+1)\ \mathrm{d}x$

Similarly, we solve:

$$\frac{-1+b}{a} = 1 \Leftrightarrow a-b=-1 \tag{50}$$

and we obtain the following expression:

$$F_3(a,b) = \begin{cases} \frac{3}{4\alpha} \\ \text{if } a=1 \text{ and } b=2 \\ \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right) \\ \text{if } a-b=-1 \text{ and } a>1 \\ \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right) \\ \text{if } a-b=-1 \text{ and } a<1 \\ 0 \text{ otherwise.} \end{cases} \quad (51)$$

Step 4:

$F_4(a,b) = a\int_{\mathbb{R}} \delta_\alpha(x-1)\delta_\alpha(ax-b-1)\,\mathrm{d}x$

We solve:

$$\frac{1+b}{a} = 1 \Leftrightarrow a-b=1 \quad (52)$$

and we have:

$$F_4(a,b) = \begin{cases} \frac{3}{4\alpha} \\ \text{if } a=1 \text{ and } b=0 \\ \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right) \\ \text{if } a-b=1 \text{ and } a>1 \\ \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right) \\ \text{if } a-b=1 \text{ and } a<1 \\ 0 \text{ otherwise.} \end{cases} \quad (53)$$

Step 5: back to the whole functional F

The study of each sub-functional has put forward the conditions for which two peaks are aligned. Putting back together the results leads to focusing on the conditions when both pairs of peaks are aligned at the same time. By comparing the conditions on a and b for each sub-functional, the only simultaneous association is between F_1 and F_4, with $a=1$ and $b=0$. In this case, we have $F(1,0) = F_1(1,0) + F_4(1,0) = \frac{3}{4\alpha} + \frac{3}{4\alpha} = \frac{3}{2\alpha}$.

Conclusion :

$$F(a,b) = \begin{cases} \frac{3}{2\alpha} \text{ if } a=1 \text{ and } b=0, \\ \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right) \\ \text{if } a>1 \text{ and } \begin{cases} a+b=\ \ 1 \\ \text{or } a+b=-1 \\ \text{or } a-b=\ \ 1 \\ \text{or } a-b=-1 \end{cases} \\ \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right) \\ \text{if } a<1 \text{ and } \begin{cases} a+b=\ \ 1 \\ \text{or } a+b=-1 \\ \text{or } a-b=\ \ 1 \\ \text{or } a-b=-1 \end{cases} \\ 0 \text{ otherwise.} \end{cases} \quad (54)$$

To conclude, we need to show that $F(a,b) \leq \frac{3}{2\alpha}$ so that the couple $(a,b)=(1,0)$ is the optimum. To this aim, we focus on: $f_1(a) = \frac{1}{2\alpha}\left(1+\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a}\right)$ for $a>1$ and $f_2(a) = \frac{1}{2\alpha}\left(a+\frac{1}{\pi(1+a)(1-a)}\sin\pi a\right)$ for $0<a<1$ and we split the proof on three parts:

- $0<a<1$
  In order to show that $f_2(a) \leq \frac{3}{2\alpha}$, we need to assess that $\frac{1}{\pi(1+a)(1-a)}\sin\pi a \leq 2$. We have:
  $$\begin{aligned} &\frac{1}{\pi(1+a)(1-a)}\sin\pi a \\ &= \frac{1}{\pi(1+a)(1-a)}\sin\pi(1-a) \\ &\leq \frac{1}{(1+a)} \leq 1 \end{aligned} \quad (55)$$
  on the interval $[0;1]$.

- $1<a<2$
  In order to show that $f_1(a) \leq \frac{3}{2\alpha}$ on the interval $[1;2]$, we need to verify that $\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a} \leq 2$. We have:
  $$\begin{aligned} &\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a} \\ &= \frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi(a-1)}{a} \\ &\leq \frac{a^3}{\pi(a+1)(a-1)} \times \frac{\pi(a-1)}{a} \\ &\leq \frac{a^2}{a+1} \leq \frac{4}{3} \end{aligned} \quad (56)$$
  on the interval $[1;2]$.

- $a>2$

$$\frac{a^3}{\pi(a+1)(a-1)}\sin\frac{\pi}{a} \leq \frac{a^2}{a^2-1} \leq \frac{4}{3} \quad (57)$$

on the interval $[2; +\infty]$.

This concludes the proof by showing that the functional has a unique global maximizer for the sought parameters $(a, b) = (1, 0)$.

References

1. Aubert, G., Kornprobst, P.: Mathematical problems in image processing: partial differential equations and the calculus of variations, vol. 147. Springer (2006)
2. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. Computer Vision and Image Understanding (CVIU) **110(3)**, 346–359 (2008)
3. Brown, L.: A survey of image registration techniques. ACM Computing Surveys **24(4)**, 325–376 (1992)
4. Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F.: SAR-SIFT: A SIFT-like algorithm for applications on sar images. In: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, pp. 3478–3481. IEEE (2012)
5. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24(6)**, 381–395 (1981)
6. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multimodal images. Medical image computing and computer-assisted intervention: MICCAI **46**, 292–299 (2006)
7. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60(2)**, 91–110 (2004)
9. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging **16(2)**, 187–198 (April 1997)
10. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27(10)**, 1615–1630 (2005)
11. Odobez, J.M., Bouthemy, P.: Robust multiresolution estimation of parametric motion models. Journal of visual communication and image representation **6**(4), 348–365 (1995)
12. Paganelli, C., Peroni, M., Pennati, F., Baroni, G., Summers, P., Bellomi, M., Riboldi, M.: Scale invariant feature transform as feature tracking method in 4D imaging: A feasibility study. In: Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, pp. 6543–6546 (2012)
13. Pluim, J., Maintz, J., Viergever, M.: Image registration by maximization of combined mutual information and gradient information. IEEE Transactions on medical imaging **19(8)**, 809–814 (2000)
14. Pratt, W.: Digital Image Processing. John Wiley and Sons, Inc., NY (1978)
15. Roshni, V., Revathy, K.: Using mutual information and cross correlation as metrics for registration of images. Journal of Theoretical & Applied Information Technology **4**(6) (2008)
16. Spontón, H., Cardelino, J.: A review of classic edege detectors. In: Image Processig On Line (2012)
17. Sun, Y., Jolly, M.P., Moura, J.F.: Integrated registration of dynamic renal perfusion MR images. In: ICIP, pp. 1923–1926 (2004)
18. Viola, P., III, W.W.: Alignment by maximization of mutual information. Proc. Vth Int. Conf. Computer Vision pp. 16–23 (June 1995)
19. Yu, L., Zhang, D., Holden, E.J.: A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. Computers & Geosciences **34**(7), 838–848 (2008)
20. Zitova, B., Flusser, J.: Image registration methods: A survey. Image and Vision Computing **21**, 977–1000 (2003)