Robust Multi-Sensor Image Alignment*

Michal Irani
Dept. of Applied Math and CS
The Weizmann Institute of Science
76100 Rehovot, Israel

P. Anandan
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052, USA

Abstract

This paper presents a method for alignment of images acquired by sensors of different modalities (e.g., EO and IR). The paper has two main contributions: (i) It identifies an appropriate image representation for multi-sensor alignment, i.e., a representation which emphasizes the common information between the two multi-sensor images, suppresses the non-common information, and is adequate for coarse-to-fine processing. (ii) It presents a new alignment technique, which applies global estimation to any choice of a local similarity measure. *In particular, it is shown that when this registration technique is applied to the chosen image representation with a local-normalized-correlation similarity measure, it provides a new multi-sensor alignment algorithm which is robust to outliers, and applies to a wide variety of* globally *complex brightness transformations between the two images.*

Our proposed image representation does not *rely on sparse image features (e.g., edge, contour, or point features). It is* continuous *and does not eliminate the detailed variations within local image regions. Our method naturally extends to coarse-to-fine processing, and applies even in situations when the multi-sensor signals are* globally *characterized by low statistical correlation.*

1 Introduction

In images acquired by sensors of *different modalities* (EO, IR, radar, etc), the relationship between the brightness values of corresponding pixels is usually complex and unknown: Visual features present in one sensor image may not appear in the other image, and vice versa; contrast reversal may occur between the two images in some image regions, while not

*Part of this work was done while the authors were at Sarnoff Corporation under NASA Ames Research Center grant NAS2-14301. It was continued while Michal Irani was at the Weizmann Institute under DARPA grant N00014-93-1-1202 through US Office of Naval Research.

in others; multiple brightness values in one image may map to a single brightness value in the other image, and vice versa. In other words, the two images are usually not correlated in their entirety, i.e., they are not *globally* correlated (often, not even statistically correlated).

There are two fundamental questions that a multi-sensor alignment algorithm should address: (i) What is a good image representation to work with (i.e., what representation will bring out the common information between the two multi-sensor images, while suppressing the non-common information)? (ii) What is an appropriate similarity measure for matching the two images within the selected representation?

Previous work on multi-sensor image alignment (e.g., [4, 6, 10, 7, 8, 11]) can broadly be classified into two major classes of algorithms. These classes differ in the way they address the two abovementioned questions:

1. Methods that use an *invariant image representation*. By invariant image representation we refer to a representation that is invariant to changes in brightness and contrast, as well as to contrast reversal. Some examples of invariant image representations are edge maps [4], oriented edge vector fields [6], contour features [7], and feature points [8]. Such representations aim at increasing the visual similarity between of the two images. Once this is achieved, registration techniques that assume similar appearance (e.g., that are based on the *brightness constancy assumption*) can be applied. For example, the registration methods employed in [4, 6] are extensions of the direct gradient-based registration methods [2, 5]).

However, in the process of creating an invariant image representation, important image information is usually lost. For example, in [4, 6, 7] there is a thresholding step. This step usually eliminates most of the detailed variations within local regions of the images, leaving only a *sparse* set of highly significant image features. Moreover, the choice of threshold is very data and sensor dependent.

2. Methods that use an *invariant similarity measure* to register the multi-sensor images, and therefore do not require an invariant image representation.

An example of such a similarity measure is *Mutual Information* [11], which is a measure of the statistical correlation between two images. The method suggested by [11] is applied directly to the raw multi-sensor intensity images, and does not require an invariant image representation. This method assumes, however, that the statistical correlation between the two images is *global*, an assumption which is often violated (e.g., Figure 4). Moreover, the statistical correlation between raw multi-sensor images tends to decrease with the reduction in spatial resolution (Section 2). Therefore, [11] in its current form does not naturally extend to coarse-to-fine estimation, which is often used to handle large misalignments. These issues will be referred to in Section 2.

In order to address the issues mentioned above, we have developed an approach which uses a *locally* invariant similarity measure while globally *constraining* the local matches. In particular, our approach to multi-sensor image alignment does *not* assume global correlation (regular or statistical) of the images, but only a local one. The underlying chosen image representation is *continuous*, and avoids thresholding and hence loss of image detail. The representation is invariant to contrast reversal, provides orientational sensitivity, and is suitable for coarse-to-fine processing. The estimation process has a built-in outlier rejection mechanism, which is critical to multi-sensor alignment due to the plurality of non-common image features across the two images (as a matter of fact, in many situations there are more "outliers" than "inliers" in a multi-sensor image pair). The motion models used in this work were 2D parametric transformations. The algorithm, however, can be extended to 3D motion models as well.

The rest of the paper is organized as follows: Section 2 describes the chosen image

representation. Section 3 describes the global alignment method with a local similarity measure. Section 4 presents results of applying our algorithm to IR/EO image pairs.

2 The Image Representation

The underlying assumption of multi-resolution alignment is that the corresponding signals at all resolution levels contain enough correlated structure to allow stable matching. This assumption is generally true when an image pair is obtained by the *same sensor*, or by two different cameras of *same modality*. However, in multi-sensor image pairs (i.e., image pairs taken by sensors of *different* modalities), the signals are correlated primarily in high resolution levels, while correlation between the signals tends to degrade substantially with the reduction in spatial resolution. This is because high resolution images capture high spatial frequency information, which corresponds to physical structure of the scene that is common to the two images. Low resolution images, on the other hand, depend heavily on illumination and on the photometric and physical imaging properties of the sensors (which are characterized by low frequency information), and these are substantially different in two multi-modality images.

To capture the common scene detail information while suppressing the non-common illumination and sensor-dependent properties, the images are transformed into high-pass *energy* images (e.g., see [3]). An example of such an energy image is a *Laplacian-energy image*, which is formed by first high-pass filtering the image with a Laplacian filter, *then squaring it*. This facilitates coarse-to-fine search based on *signal details*. In [3] the Laplacian-energy image is used for effectively detecting small (high-resolution) temporal changes already at low resolution levels.

High-pass energy image representations are useful for multi-sensor alignment, because:

(i) The creation of such energy images does not involve any thresholding, and therefore preserves all image detail. This is in contrast to "invariant" representations (e.g., edge maps [4], edge vectors [6], contours [7], point features [8]), which eliminate most of the detailed variations within local image regions.

(ii) The image information which is eliminated in the creation of the high-pass energy images is exactly that which is *not* common to the two multi-sensor images. In particular: (a) the sensor-dependent low-resolution information is eliminated, and (b) contrast-reversal which may occur between the sensors (e.g., Fig. 3) is removed by the squaring operation. In other words, the energy image representation is *invariant to contrast reversal*.

(iii) As mentioned in [3], a pyramid data structure of the high-pass energy image projects high resolution signal information into low resolution levels. In our case, this facilitates coarse-to-fine alignment based on *correlated* scene details, as opposed to using pyramids of the raw multi-sensor images (which contain *uncorrelated* sensor information at low spatial resolutions).

However, the Laplacian, being a rotationally invariant operator, does not preserve directional information. This leads to potential false correspondences of patterns that are oriented along different directions in the Laplacian energy images. The energy-image representation that we use is based on *directional-derivative* filters rather than a Laplacian filter. On top of the abovementioned advantages of high-pass energy images, the *directional-derivative*-energy also preserve *directional* information, and thereby avoid this problem. This further enhances the robustness of the registration algorithm against the numerous outliers so common in a multi-sensor image pair.

The directional derivative filter is applied to the raw image in four directions (horizontal, vertical, and the two diagonals). Then, each of the four generated derivative images is squared. (Since the squaring operation doubles the frequency band, the raw image is filtered

with a Gaussian prior to the derivative filtering, to avoid aliasing effects).

The alignment algorithm (Section 3) is applied *simultaneously* to all 4 corresponding multi-sensor pairs of directional-derivative-energy images, seeking a *single* parametric transformation $\vec{p}$, which *simultaneously* brings all directional pairs into alignment (see Section 3).

Fig. 1 shows an example of the four directional-derivative-energy pairs constructed from a multi-sensor image pair.

Fig. 2 shows the Gaussian pyramid constructed for one of the four multi-sensor *pairs* of directional-derivative-energy images.

3 The Alignment Algorithm

To align the multi-band energy image representation (Section 2), our alignment algorithm uses a *local* correlation-based similarity measure, *without* assuming global correlation (regular or statistical) between the images. We have applied the algorithm with a normalized-correlation-based local similarity measure for reasons explained below. However, it can be similarly applied with a *local* statistical-correlation-based similarity measure (e.g., based on Mutual Information), or *any other appropriate local measure*.

The global parametric estimation is applied *directly* to the collection of all local *correlation surfaces*, while avoiding an independent local search for peaks in the individual surfaces. Global alignment has the advantage of directly estimating the global parametric transformation, without first committing to any particular matches locally. In other words, local matching is constrained by global alignment. Such a scheme is useful in any alignment algorithm, but is particularly critical in multi-sensor alignment, due to the plurality of outliers across sensors and hence the unreliability of local matches. Although global alignment has been used for image registration, it has been based on minimizing the intensity differences

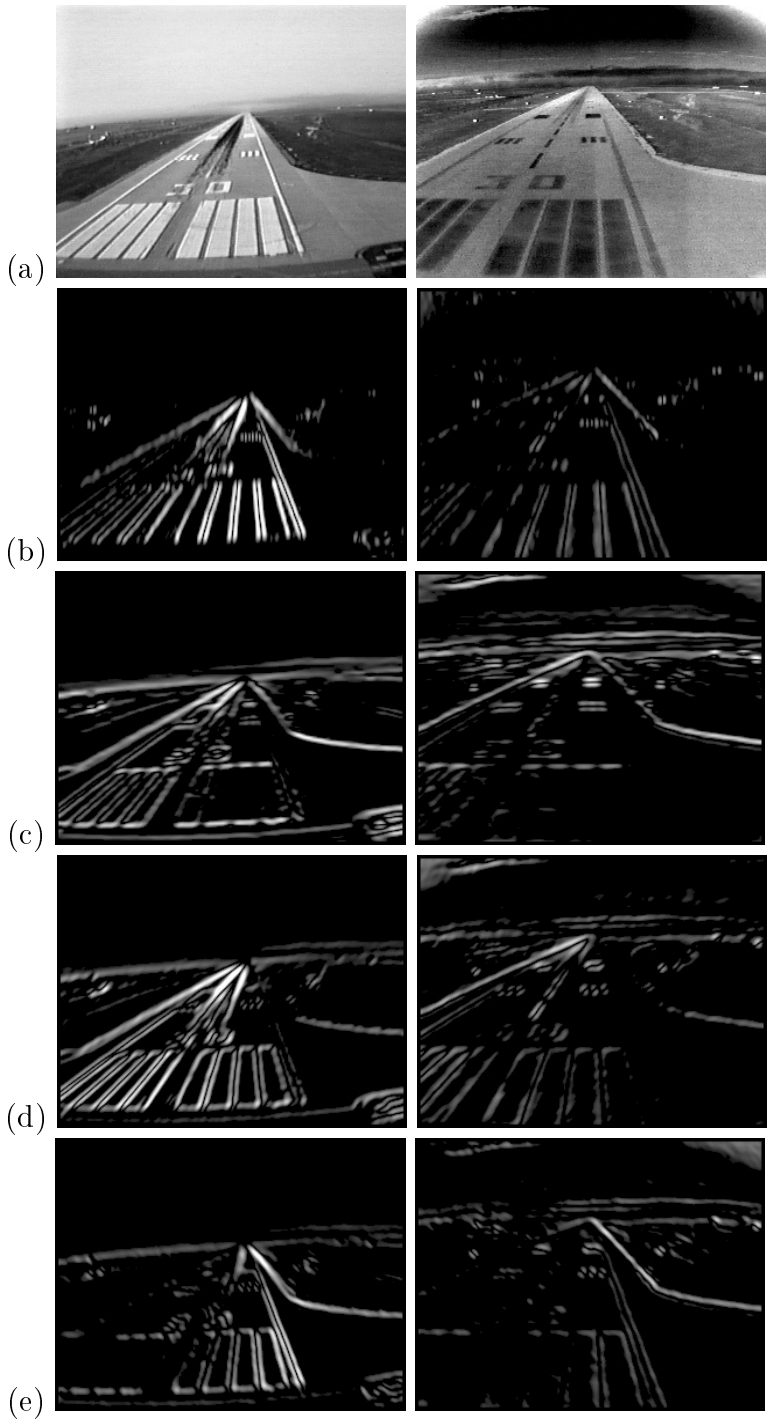

Figure 1: The four directional-derivative-energy image pairs. *Left column:* EO. *Right column:* IR.
(a) The raw multi-sensor image pair. (b) horizontal derivative energy, (c) vertical derivative energy, (d,e) energies of diagonal derivatives.

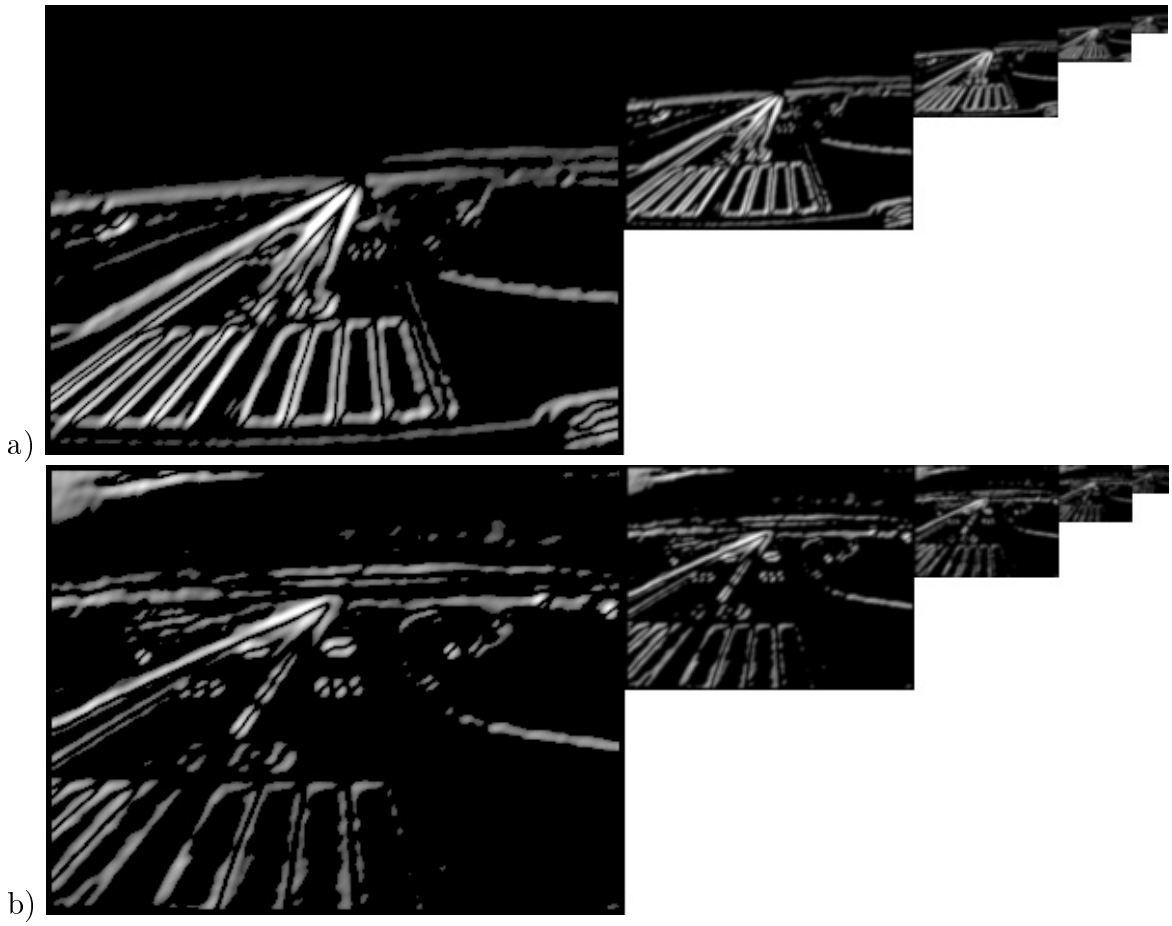

Figure 2: The Gaussian pyramid constructed for one of the four pairs of directional-derivative-energy images (Fig. 1.d): (a) EO. (b) IR.

between the corresponding pixels in the two images, i.e., using the sum of squared differences (SSD) as the match measure. That is, the similarity measure is based on the "brightness constancy" assumption, which is severely violated in a multi-sensor image pair (even in the energy images). In this work, we have *generalized* global alignment techniques to use *any local similarity measure* (e.g., normalized correlation, SSD, or any other measure) which is suitable for the particular alignment problem. This is done via global regression applied *directly* to the local *similarity-measure surfaces* (e.g., correlation surfaces), as described in Section 3.1. In particular, we found normalized-correlation to be a suitable similarity measure for multi-sensor energy-image alignment.

Global alignment is particularly critical when using the directional-derivative images: no prior local estimation process can produce meaningful local matches on a directional-derivative image pair, as these images lack information in the direction *perpendicular* to the directional derivative (the "aperture problem"). The *simultaneous* and *global* registration of all (four) directional pairs, however, provides full directional information.

Motion Models: When the scene can be approximated by a planar surface, or when the baseline between the two sensors is small relative to their distance from the scene, then the displacement field between the two images can be modeled in terms of a single 2D parametric transformation (see [2] for a taxonomy of motion models).

We have focused our attention on alignment using a 2D parametric transformation, although our approach generalizes to 3D models as well. Specifically, we focus on parametric transformations which are *linear* in their unknown parameters $\{p_i\}$. For such transformations, the motion vector $\vec{u}(x,y) = (u(x,y), v(x,y))^T$ can be expressed as:

$$\vec{u}(x,y;\vec{p}) = X(x,y) \cdot \vec{p}, \tag{1}$$

where $X(x,y)$ is a matrix which depends only on the pixel coordinates (x,y), and $\vec{p} = (p_1, ..., p_n)^T$ is the unknown parameter vector. For example, for an *affine* transformation:

$$\begin{bmatrix} u(x,y;\vec{p}) \\ v(x,y;\vec{p}) \end{bmatrix} = \begin{bmatrix} p_1 + p_2 x + p_3 y \\ p_4 + p_5 x + p_6 y \end{bmatrix}, \tag{2}$$

therefore, in this case: $\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_6)^T$ and

$$X = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix},$$

and for a *quadratic* transformation:

$$\begin{bmatrix} u(x,y;\vec{p}) \\ v(x,y;\vec{p}) \end{bmatrix} = \begin{bmatrix} p_1 + p_2 x + p_3 y + p_7 x^2 + p_8 xy \\ p_4 + p_5 x + p_6 y + p_7 xy + p_8 x^2 \end{bmatrix},$$

therefore: $\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)^T$ and

$$X = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 \end{bmatrix}.$$

The Normalized-Correlation as a Local Similarity Measure: Normalized-correlation of two signals is invariant to local changes in mean and contrast. In other words, when the two signals are *linearly related*, their normalized-correlation is 1. When the linear relationship does not hold, but the two signals contain *similar spatial variations* (as measured in the form of local fluctuations), the normalized-correlation will still give a value close to unity.

In general, however, the global relationship between two multi-sensor images is complex, and therefore the two signals are not globally correlated (even after computing the energy images). Statistical correlation is a better *global* measure than regular or normalized correlation, but may still not be a strong enough global similarity measure, because multiple brightness values in one image may map to a single brightness value in the other image, and vice versa. *Locally*, however, within *small* image patches which contain corresponding image features, statistical correlation is high. Normalized-correlation is a linear approximation of the statistical correlation of two signals in a small window, and is cheaper to compute.

The *energy* images that we compute tend to highlight the local variations that correspond to local structure in the scene. These images are invariant to contrast reversal, but vary in mean and contrast. When the relationship between corresponding patches deviates from linear, the normalized-correlation (applied over local windows) is less than 1, but is still high for the correct displacement. For other displacements the normalized-correlation will be low, especially for highly textured image patches. Therefore, the local normalized-correlation surface of such patches will be concave with a prominent peak at the correct displacement.

For corresponding image patches that contain *mutually exclusive* image features (i.e., image features which appear in only one of the 2 multi-sensor images – a thing which occurs frequently), the local correlation surface will not have a concave shape with a prominent peak. Therefore, the *structure* of the local-normalized-correlation *surfaces* provides useful information for alignment. The information from all of these local structures, however, should be *simultaneously* used to determine the global alignment parameters. This is essential to avoid the numerous potential false matches in limited local analysis. This is achieved via global regression applied *directly* to the collection of local normalized-correlation surfaces, as described in Section 3.1.

3.1 Global Alignment with Local Correlation

Given two images, f and g, and their directional-derivative energy images, $\{f_i\}_{i=1}^4$ and $\{g_i\}_{i=1}^4$, find the parametric transformation $\vec{p}$ which maximizes the sum of all local normalized-correlation values. Let $S_i^{(x,y)}(u,v)$ denote a correlation *surface* corresponding to a pixel (x,y) in f_i. For any shift (u,v) of g_i relative to f_i, $S_i^{(x,y)}$ is defined as:

$$S_i^{(x,y)}(u,v) \stackrel{\text{def}}{=} f_i(x,y) \circ_{\mathbf{N}} g_i(x+u, y+v)$$

where $\circ_{\mathbf{N}}$ denotes normalized correlation computed over a small window. Let $\vec{u} = (u(x,y;\vec{p}), v(x,y;\vec{p}))$ denote the motion field described by the parametric transformation $\vec{p}$. Then the parametric registration problem can be stated as follows: Find the parametric transformation $\vec{p}$ that maximizes the *global* similarity-measure $M(\vec{p})$:

$$\begin{aligned} M(\vec{p}) &= \Sigma_{x,y}\Sigma_i S_i^{(x,y)}(u(x,y;\vec{p}), v(x,y;\vec{p})) \\ &= \Sigma_{x,y}\Sigma_i S_i^{(x,y)}(\vec{u}(x,y;\vec{p})). \end{aligned} \tag{3}$$

To solve for $\vec{p}$ that maximizes $M(\vec{p})$, we use Newton's method [9], which iteratively fits quadratic approximations to the objective function, and refines the peak location that maximizes these quadratic surfaces. In order to provide the context for our use of Newton's method for the maximization problem at hand, we first briefly outline the steps of this method.

Given the current estimate of the motion parameters $\vec{p_0}$, let

$$M(\vec{p}) = M(\vec{p_0}) + (\nabla_{\vec{p}} M(\vec{p_0}))^T \vec{\delta_p} + \vec{\delta_p}^T H_M(\vec{p_0}) \vec{\delta_p} \tag{4}$$

denote the quadratic approximation of $M(\vec{p})$ around $\vec{p_0}$, where, $\vec{\delta_p} = \vec{p} - \vec{p_0}$ is the unknown refinement step of $\vec{p_0}$ that we want to solve for, $\nabla_{\vec{p}} M$ denotes the gradient of M, and H_M denotes the Hessian of M (i.e., the matrix of second derivatives), both computed around $\vec{p_0}$. According to Newton's method [9], the refinement $\vec{\delta_p}$ computed based on this approximation is:

$$\vec{\delta_p}^* = -(H_M(\vec{p_0}))^{-1} \cdot \nabla_{\vec{p}} M(\vec{p_0}) \tag{5}$$

To apply the Newton's refinement step to our problem, we derived the expressions for $\nabla_{\vec{p}} M$ and H_M *in terms of the measurable image quantities*, i.e., the collection of correlation surfaces $\{S_i^{(x,y)}\}$: Using the chain-rule of differentiation, we obtain

$$\begin{aligned} \nabla_{\vec{p}} M(\vec{p}) &= \Sigma_{x,y,i} \nabla_{\vec{p}} S_i(\vec{u}) = \Sigma_{x,y,i} (X^T \cdot \nabla_{\vec{u}} S_i(\vec{u})) \\ H_M(\vec{p}) &= \Sigma_{x,y,i} (X^T \cdot H_{S_i}(\vec{u}) \cdot X) \end{aligned} \tag{6}$$

where X is the matrix defined in Eq. (1), $\nabla_{\vec{u}} S_i$ is the gradient of $S_i^{(x,y)}(\vec{u})$, and H_{S_i} is the Hessian of $S_i^{(x,y)}(\vec{u})$.

In other words, the quadratic approximation of M around $\vec{p_0}$ is obtained by combining the quadratic approximations of each of the local correlation surfaces $\{S_i^{(x,y)}\}_{x,y,i}$ around the

local displacement vector $\vec{u_0} = \vec{u}(x, y; \vec{p_0})$, which is induced at pixel (x, y) by the parametric transformation $\vec{p_0}$ (estimated at the previous iteration).

Substituting Eqs. (6) into Eq. (5) provides an expression for the refinement step $\vec{\delta_p}^*$ in terms of the correlation surfaces $\{S_i^{(x,y)}\}$:

$$\vec{\delta_p}^* = -(\Sigma_{x,y,i}\ X^T H_{S_i}(\vec{u_0}) X)^{-1}\ (\Sigma_{x,y,i}\ X^T \nabla_{\vec{u}} S_i(\vec{u_0})) \tag{7}$$

Note that these steps do not make any assumptions about the local correlation surface, except that it is twice differentiable. Thus, *any local similarity-measure* can be substituted for correlation, and our method will still apply.

The steps of the algorithm: To account for large misalignments between pairs of images, we perform multi-resolution coarse-to-fine estimation, e.g., as in [2]. A Laplacian (or a Gaussian) pyramid is constructed for each of the energy images. Let f_{il} and g_{il} $(i = 1, 2, 3, 4)$ denote the directional-derivative energy images at resolution level l in the pyramids of f_i and g_i, respectively. Starting at the coarsest resolution level with $\vec{p_0}$ initially set to 0, the following steps are performed at each resolution level:

1. For each pixel (x, y) at f_{il} $(i = 1, 2, 3, 4)$, compute a local normalized-correlation surface around the displacement $\vec{u_0}$ (i.e., around the displacement estimated at the previous iteration). In practice, the correlation surface is estimated only for a small number of displacements $\vec{u}$ of g_{il} *within a radius* d around $\vec{u_0}$, i.e.:

$$S_{li}^{(x,y)}(\vec{u}) = f_{il}(x, y) \circ_{\mathbf{N}} g_{il}(x + u, y + v)\ ,$$

$$\forall \vec{u} = (u, v)\ s.t.\ ||\vec{u} - \vec{u_0}|| \leq d$$

where the radius d is determined by the size of the masks used for discretely estimating the

first and second order derivatives of $S_{l_i}^{(x,y)}(\vec{u})$ at $\vec{u_0}$. In our current implementation we used Beaudet's masks [1] to estimate the first and second order derivatives of the surfaces. We have experimented both with 3×3 masks (i.e., $d = 1$) and with 5×5 masks (i.e., $d = 2$).

2. Perform the regression step of Eq. (7) to compute the parametric refinement $\vec{\delta_p}^*$.

3. Update $\vec{p_0}$: $\vec{p_0} := \vec{p_0} + \vec{\delta_p}^*$, and go back to step 1.

After repeating the above process for a few iterations (typically 4), the parameters $\vec{p}$ are propagated to the next resolution level, and the process is repeated at that resolution level. The process is stopped when the iterative process at the highest resolution level is completed.

In practice, to improve performance, we add an *image warping* step before each iteration (as in [2]). The inspection images $\{g_i\}$ are warped towards the reference images $\{f_i\}$ according to the current estimated parametric transformation $\vec{p_0}$. After warping the images, $\vec{p_0}$ is set to 0, and $\vec{\delta_p}^*$ is estimated between the pairs of references and warped inspection images. Warping compensates for the spatial distortions between the pairs of images (e.g., scale difference, rotations, etc), and hence improves the quality of the correlation.

Outlier rejection: To further condition and robustify the regression step of Eq. (7), only pixels (x, y) for which the quadratic approximation of $S_i^{(x,y)}(\vec{u})$ around $\vec{u_0}$ is *concave* are used in the regression process. Other pixels are ignored. Since corresponding multi-sensor image patches which have *mutually exclusive* image features will not tend to have a concaved-shaped local correlation surfaces, they will be eliminated from the regression at this point. Moreover, the contribution of each pixel to the regression step is weighted

by the determinant of it Hessian. This built-in *outlier rejection* mechanism provides the algorithm with a strong *locking* property onto a dominant parametric motion, even in the presence of independent motions, noise, and exclusive features that appear in only one of the sensor-images (but not in the other).

4 Examples

The alignment algorithm described in Section 3 was implemented and applied with an affine parametric model (Eq. 2) to pairs of multi-sensor images. Fig. 3 shows result of alignment of two multi-sensor images (visible and IR) obtained by sensors mounted on an aircraft approaching landing. Note the significant difference in scale between the two images (due to significantly different internal sensor parameters). Also note that contrast reversal occurs in some parts of the images (e.g., runway markings), while not in others (e.g., runway boundaries).

The algorithms has been applied successfully even in very challenging situations, such as the one shown in Fig 4. Note the significant difference in image content between the two sensor-images. Apart from having significantly different appearance, there are many non-common features (i.e., *outliers*) in the multi-sensor image pair, which can theoretically lead to false matches. These are overcome by the built-in outlier mechanism of our algorithm (see Section 3).

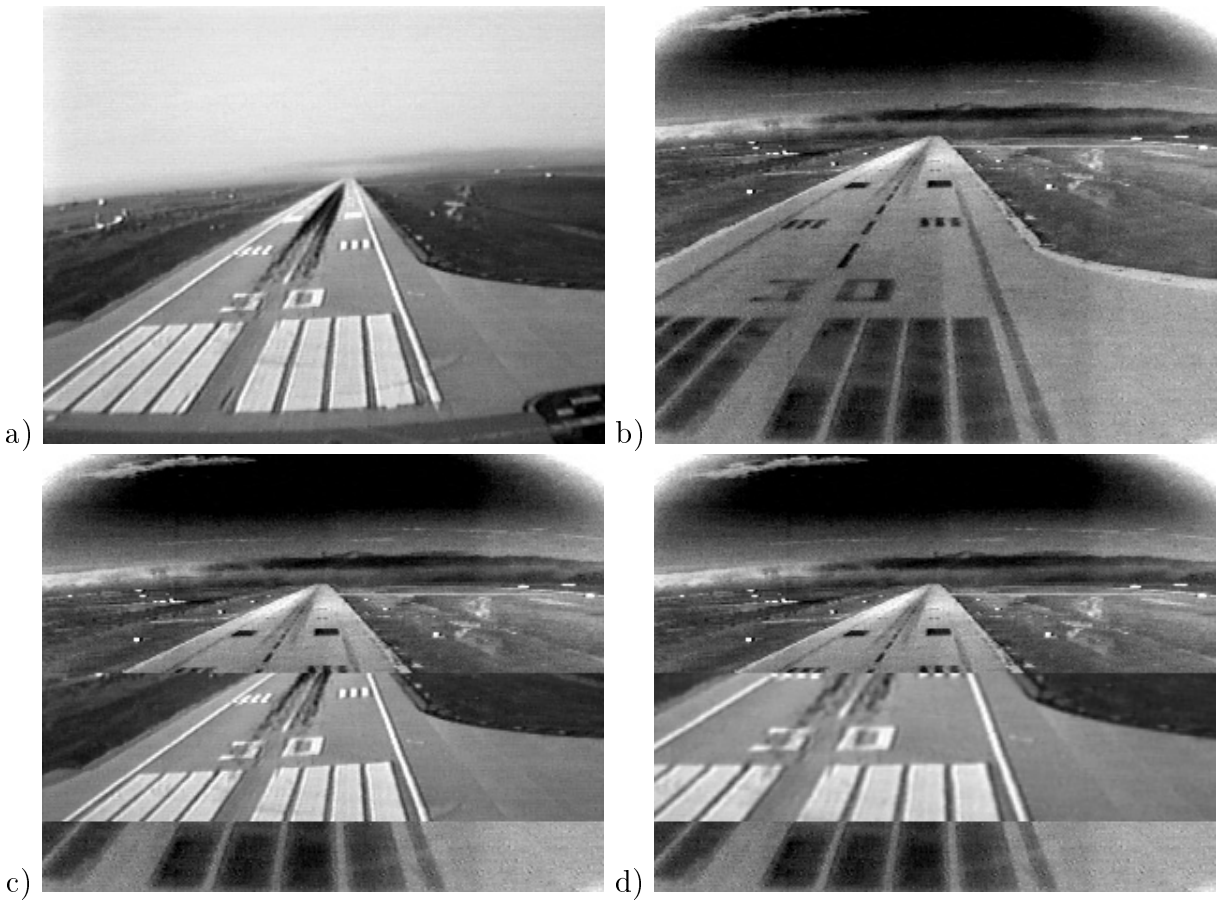

Figure 3: Multi-sensor Alignment.
(a) EO image. (b) IR image. (c) Composite display of the two multi-sensor images *before* alignment. Horizontal strips from the two images are spliced together. Note the significant misalignments between the images (e.g., the runway markings and the borders of the runway). (d) Composite (spliced) display of the two multi-sensor images *after* alignment. Note that all structures in the scene are aligned.

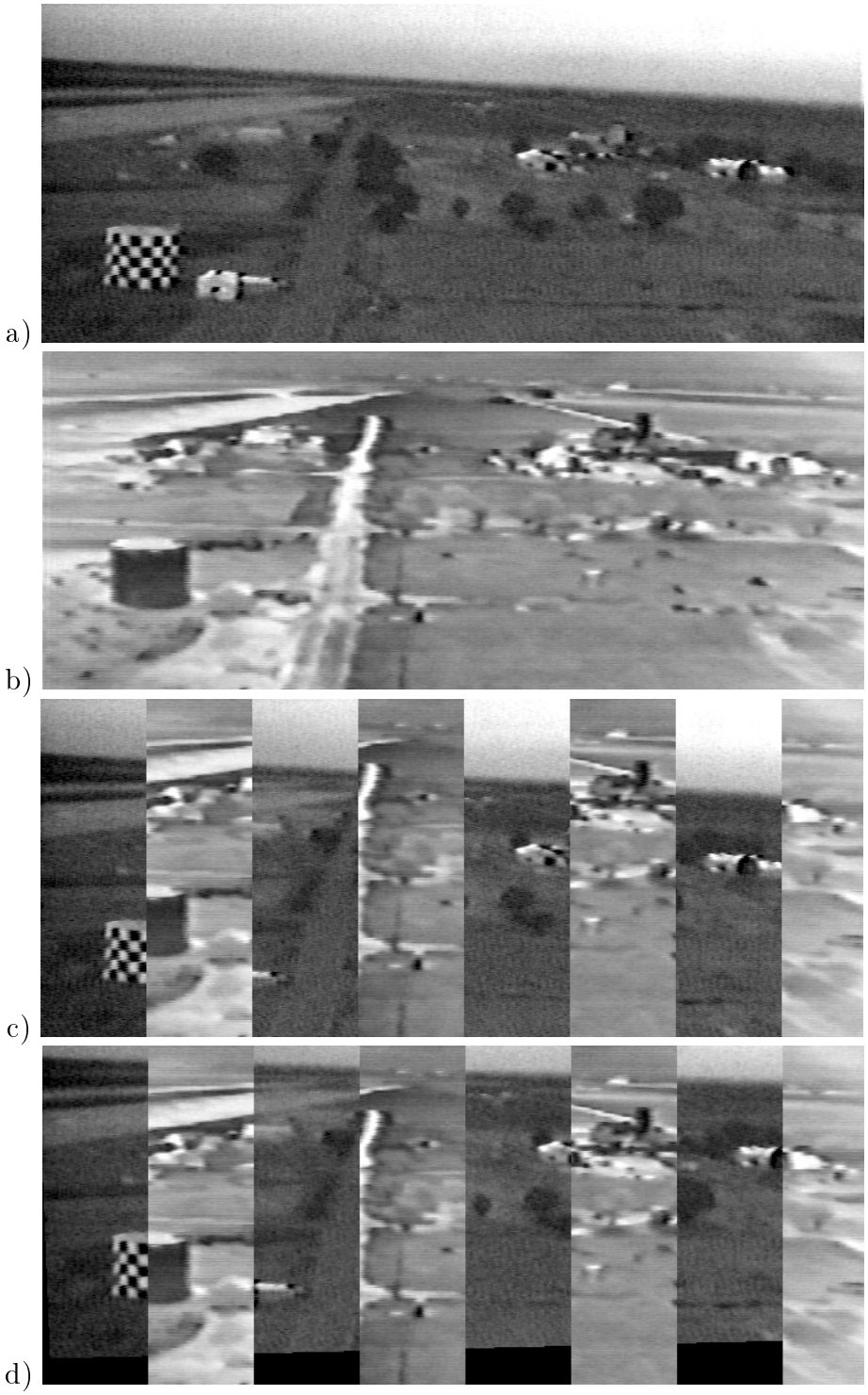

Figure 4: Multi-sensor Alignment.
(a) EO image. (b) IR image. (c) Composite (spliced) display *before* alignment. (d) Composite (spliced) display *after* alignment. Note in particular the perfect alignment of the water-tank at the bottom left of the images, the building with the arched-doorway at the right, and the roads at the top left of the images.

Acknowledgement

The authors would like to thank Peter Burt for his inspiration and guidance during this work.

References

[1] Paul R. Beaudet. Rotationally invariant image operators. In *International Conference on Pattern Recognition*, pages 579–583, 1978.

[2] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision*, pages 237–252, Santa Margarita Ligure, May 1992.

[3] P.J. Burt. Smart sensing with a pyramid vision machine. *Proceedings of the IEEE*, 76:1006–1015, 1988.

[4] K. J. Dana and P. Anandan. Registration of visible and infrared images. In *Proc. SPIE Conf. on Arch., Hardware and FLIR in Auto. Targ. Rec.*, pages 1–12, 1993.

[5] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1):5–16, January 1994.

[6] R. Kumar, K. Dana, and P. Anandan. Frameless registeration of mr and ct 3d volumetric data sets. In *Proc. of the Workshop on Applications of Computer Vision II*, Sarasota, Fl., 1994.

[7] H. Li, B.S. Manjunath, and S.K. Mitra. A contour-based approach to multisensor image registration. *IEEE Trans. on Image Processing*, pages 320–334, 1995.

[8] H. Li and Y.T. Zhou. Automatic eo/ir sensor image registration. In *IEEE Int. Conf. on Image Proc.*, volume B, pages 161–164, 1995.

[9] David G. Luenberger. *Linear and Nonlinear Programming.* Addison Wesley, 1984.

[10] P. A. van den Elsen and M.A. Viergever. Marker-guided multimodality matching of the brain. *European Radiology*, 4(1):45–51, 1994.

[11] P. Viola and W. Wells III. Alignment by maximization of mutual information. In *International Conference on Computer Vision*, pages 16–23, Cambridge, MA, June 1995.