Real-world multisensor image alignment using edge focusing and Hausdorff distances

Yunlong Sheng[*a], Xiangjie Yang[a], Daniel McReynolds[a], Zhong Zhang[a], Langis Gagnon[b] and Léandre Sévigny[c]

[a] Université Laval, Dept. de Physique, Ste-Foy, Qc., Canada G1K 7P4
[b] Lockheed Martin Canada, 6111 Royalmount Ave.,Montréal, Qc, Canada, H4P 1K6[§]
[c] Defence Research Establishment Valcartier, 2459 Boul. Pie XI nord, C.P. 8800, Courcelette, Qc, Canada G0A 1R0

Keywords: Image registration, Feature inconsistency, Edge focusing, Saliency, Hausdorff distance

ABSTRACT

The area-based methods, such as that using the Laplacian pyramid and Fourier transform-based phase matching, benefit by highlighting high spatial frequencies to reduce sensitivity to the feature inconsistency problem in the multisensor image registration. The feature extraction and matching methods are more powerful and versatile to process poor quality IR images. We implement multi-scale hierarchical edge detection and edge focusing and introduce a new salience measure for the horizon, for multisensor image registration. The common features extracted from images of two modalities can be still different in detail. Therefore, the transformation space match methods with the Hausdorff distance measure is more suitable than the direct feature matching methods. We have introduced image quadtree partition technique to the Hausdorff distance matching, that dramatically reduces the size of the search space. Image registration of real world visible/IR images of battle fields is shown.

1. INTRODUCTION

Most advanced vision systems utilize multiple imaging sensors to capture more information. The multi-spectrum sensors, imaging spectrometers and lateral synthetic aperture radar are widely used for remote sensing. Computer tomography, nuclear magnetic resonance imaging and ultrasound scanning imaging provide multiple modality images for computer aided diagnosis in the biomedical applications.

We are interested in fusion of two real-world image sequences taken with two broad band cameras mounted on a ground vehicle. One is in visible and another is in infrared (IR) band. Image registration is necessary before fusion. Unlike the problems of multisensor image registration in remote sensing and imaging spectrometry, where the electromagnetic spectral responses of the sensors vary smoothly from one sensor to another, our data are from two well separated spectral bands, so that the radiometric data disparities are significant, and feature inconsistency in two image modalities is important. The technical challenge is then to extract the structurally salient edge features, which are common in the visible and IR images, and to use them for image registration, knowing that the same edges extracted from two different modality real-world images can still have different details, so that exact match between them is impossible.

In this paper we analyze the feature inconsistency problem and review the area-based approaches for multisensor image registration, and we implement the feature based approach. We use multi-scale hierarchical edge detection and edge focusing

[*] Correspondence: Email sheng@phy.ulaval.ca; WWW: http://fourier.phy.ulaval.ca/sheng/Yunlong-Sheng.htm Phone: 418-656-3908, Fax. 418-656-2623

§ Current address: Centre de Recherche Informatique de Montreal, 550 Sherbrooke Ouest, Montreal (Quebec), CANADA, H3A 1B9.

and the edge salience measure to extract salient edges from the low contrast and noisy IR image background. We use the Hausdorff distance measure for an
optimal match between the extracted curves from two different modalities. We introduce the image partitioning technique in the Hausdorff distance matching, so that the affine transformation is approximated by local translations. This speeds up the Hausdorff distance matching process.

2. FEATURE INCONSISTENCY IN VISIBLE AND IR IMAGES

Multiple sensors have different electromagnetic spectral responses to capture distinguished features in different spectral bands. One of the principles in the design of multi-sensor imaging systems is to maximize the independence of the acquired data. One selects imaging wavelengths in an electromagnetic spectral band as wide as possible. In principle, the images from multiple sensors should be uncorrelated and independent from each other. This is natural, since if one sensor captures images that are similar or correlated to the images already obtained by another sensor, then this sensor provides no additional useful information and can be withdrawn from the system.

Our task is to register two broad band images from visible and infrared bands. The images from different bands of the electromagnetic spectrum have different radiometric properties. The IR passive sensors have radiometric data, which consist of 1) energy emitted by thermal radiation from the object bodies; 2) atmospheric emission reflected from object surfaces. When the reflected solar or lunar radiation is absent, the night eye would respond only to photons emitted by the body temperature. The thermal images for night vision are produced primarily by self-emission and by emissivity differences among the objects on the scene. All of the scene temperature, emissivity and reflectivity contributions taken together can be represented at any point in the scene by an effective temperature at that point. The effective temperature is that at which a blackbody radiator would produce the measured irradiance near the point. In the general case, the gray-scale level of the IR image would depend on the difference in temperature, emissivity and reflectivity of the objects in the scene.

There are significant gray-level disparities between the IR and visible image. The thermal emitters are not necessarily good visual reflectors. A surface of high visual reflectivity (white surface) in visible band usually has low emissivity, so that those visually bright objects in the visible image may be dark in the thermal scene and vice versa. The sky is usually the brightest region in the visible image. It is, however, a dark region in the IR image because of the low temperature of the background and the lack of reflectance. This is the reversal of contrast polarity between the visible and IR images.

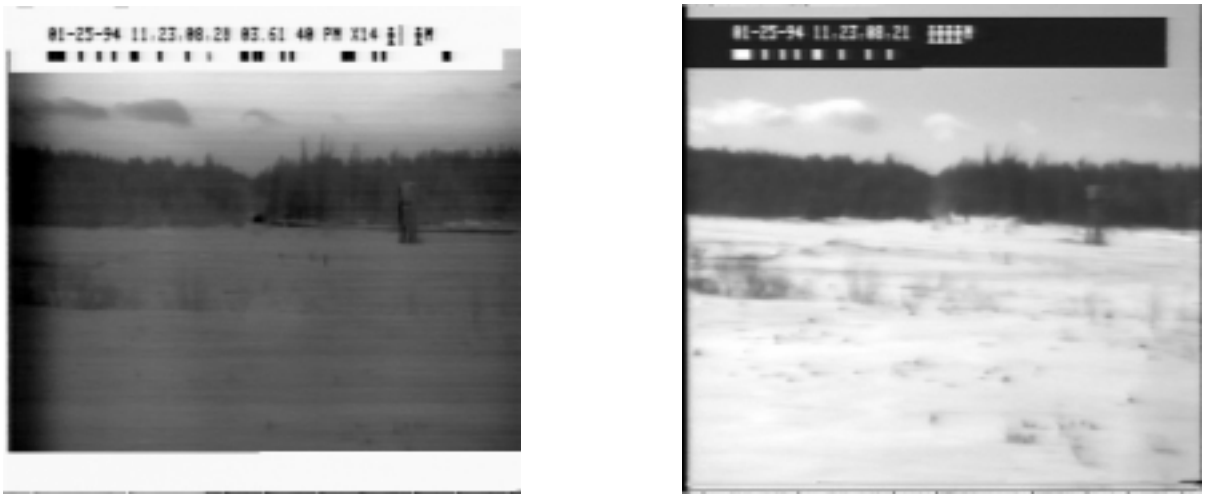


Fig.1 Contrast reversed IR image (left) and visible image (right)

The real-world natural out-door images have more complicated grayscale level disparities and contrast polarity reversals. Fig. 1 shows a contrast reversed IR image, obtained from the popular "XV" application, compared with the corresponding visible image of the same scene. In most regions, the IR and visible images have reversed contrast polarity. However, some regions,

such as the clouds in the sky, can have the same polarity of contrast in both IR and visible images. The clouds in the sky are brighter than the sky background in the visible image because of its higher reflectivity. The clouds are also brighter than the sky in the original IR image because of its higher reflectivity and emissivity. In the contrast polarity reversed IR image, the clouds in the sky becomes a region darker than the sky. Apart from the sky and clouds there are still important gray-level disparities between the contrast polarity reversed IR image and the visible image in Fig.1. In the mathematical sense the two images are not similar at all. Hence, a simple reversal of contrast polarity can not remove all the gray level disparities. Also, shadows in the visible images are absent in the IR images.

3. AREA-BASED REGISTRATION

The area-based approach and the optical flow approach for the image registration usually require the radiometric data of two images to be similar. If these distributions are very different, then area-based matches will fail. However, the area-based approaches have advantages for the use of full image information, so that they may be applied to images with rich or poor structure, and for the robustness against random noise. The area-based methods can be still used for multisensor image registration as described in the following.

3.1 Laplacian pyramid

There are several approaches that address the registrability problem. One is to transform the two images whose radiometric intensity distributions are dissimilar into two images that have similar intensity distributions. This transformation can be the nonlinear transform of the Laplacian pyramid coefficients. The intensity of the Laplacian pyramid images is insensitive to gray-scale level disparities and polarity reversals of contrast. Therefore, the Laplacian pyramid representation can be used for area-based multisensor image registration.

Burt and Adelson introduced the Laplacian pyramid for multiresolution image decomposition, coding and reconstruction[1]. The original image is first averaged in neighbouring pixels by a convolution with the low-pass Gaussian filter. This averaging reduces the resolution of the image. Hence, a down-sampling by a factor of two on the averaged image can be applied, resulting in the averaged approximation of the input image. The averaging and down-sampling process can be iterated with the smoothing filter dilated by a scale factor of 2 in each resolution level. Hence, the original data is represented by a set of successive approximations. Each approximation corresponds to a smoothed version of the original data at a given resolution, constituting a Gaussian pyramid.

The high frequency details in the image are lost in the Gaussian pyramid by the low-pass filtering and down sampling. One has to compute the difference between the averaged image and the original image in two successive pyramid levels. These difference images contain the detail information of the image. All the difference images form a new set of sequences that constitute another pyramid called the Laplacian pyramid. The original signal can be reconstructed exactly by summing the Laplacian pyramid.

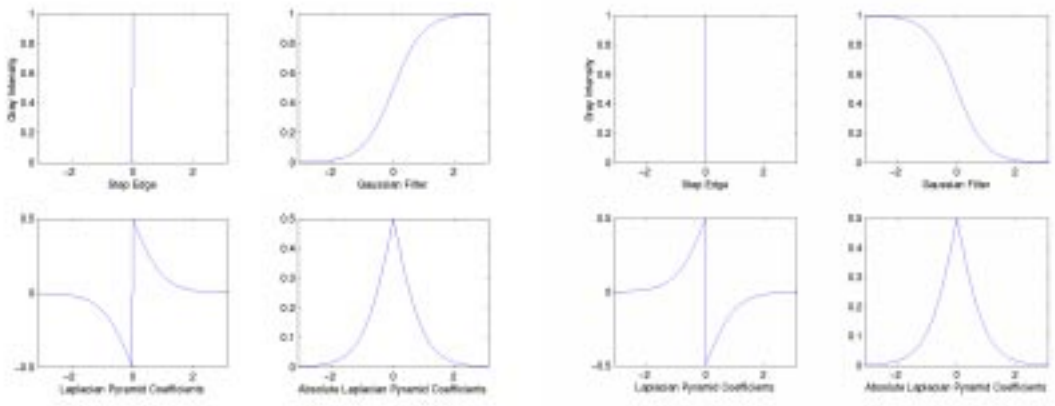


Fig.2 Two step edges with opposite polarities, smoothed by Gaussian filter, their Laplacian pyramid and absolute coefficients of the Laplacian pyramid.

Fig. 2 shows two step edges with opposite polarities of contrast. The edges are smoothed by the Gaussian filter. The Laplacian pyramid coefficient is the difference between the original edge and the smoothed edge. This is the Laplacian

pyramid image at a resolution level. When we take the absolute values of the Laplacian pyramid coefficients, the two Laplacian pyramid images become the same for the two contrast reversed step edges. Then, the area-based image registration can be applied to the Laplacian pyramid image intensities[2].

3.2 Phase matching in the Fourier transform-based registration

Images from different sensors have different radiometric intensity distributions due to the different spectral responses of the sensors. Those differences appear mostly as slow variations over wide regions in the image, such as sky, land and forest, which are usually represented with low spatial frequencies and are concentrated in a narrow low frequency band.

In the Fourier transform-based registration[3], the displacement is found by cross-correlation between two images. The location of the cross-correlation peak mainly depends on the Fourier spectrum phase and is insensitive to Fourier spectrum energy. One can then whiten the Fourier spectrum and use only the spectrum phase information. In this approach, the phases in the low and high frequency spectral bands contribute equally to the cross-correlation. Therefore, the contribution of the high frequencies is greatly highlighted. The location of the cross-correlation peak would not change if the image intensity variations are limited to a narrow spatial frequency band. The Fourier phase correlation registration method is then relatively independent of the sensors.

3.3 Block matching

The area-based image registration using cross-correlation can account only for image translations. To fit more general image transformations such as affine transformation, one can partition images into a number of sub-images, which are located in a regular grid, and then, define a central window in each sub-image as a template and correlate those blocks with another image. The template matching results in a number of displacement vectors, which are evenly distributed over the image. The use of the cross-correlation for the block matching implies that the sub-images are only translated in a small neighborhood of the respective grid, although the images can have more complicated distortions. The local translations of the blocks describe approximately the global affine or even local distortions. Therefore, the local displacement vectors can be used to fit more complex image distortion model[4].

In the multisensor image registration the image grayscale level disparities can not be removed completely by a reversal of contrast polarity, as shown in Fig.1, by the Laplacian pyramid representation or by whitening the Fourier spectrum. The residual disparities in some blocks can lead to erroneous displacement vectors, which will be the outliers for fitting to image transformation. The block matching approach is more appropriate in this case[5], since by using the robust image fitting techniques, such as the least median of squares or M-estimation, this approach can be robust against outliers.

4. FEATURE-BASED REGISTRATION

We notice that both the Laplacian pyramid representation and the phase matching technique benefit from the use of high spatial frequencies of the image. The values of the Laplacian pyramid image are equal to zero in the regions where image intensity is uniform. These values are small in the regions of slow image intensity variation. The Laplacian pyramid therefore represents detailed information, namely contours, in the image. In the phase matching approach the whitening of the Fourier spectrum highlights the high spatial frequencies. The inverse Fourier transform of the whitened spectrum is an edge and contour enhanced image. In some sense both the Laplacian pyramid representation and the phase matching share the same features with the edge matching approaches for image registration. However, the edge extraction by the Laplacian pyramid and by the whitening Fourier spectrum is not as powerful and precise as that implemented by using directly the edge detectors. Especially, for our real-world IR images, which are typically of low contrast and noisy, the Laplacian pyramid approach is not able to extract salient contours for image matching, so that feature-based image registration is adopted.

The feature-based approach requires extracting the common features from two images. If both types of images represent the same real world objects, then the objects should appear in both image types. While images may appear differently in different sensors, different objects appear always differently in the multisensor images, no matter what is the spectral response of the imaging sensors. As a result, the boundaries between objects would be preserved, although we find that the edges extracted from the same real world objects in two image types still can have different details due to the differences in the radiometric responses of two sensors.

Advantages of the feature-based image registration are that the common image edge features are not sensitive to the spectral responses of the multiple sensors; the processing speed is independent of image displacement; any image transformations can be accounted for and the powerful and versatile edge detection, edge saliency techniques can be used.

5. MULTI-SCALE EDGE DETECTION

In the 3-D real world scene, objects are separated from the background by depth discontinuities, which are usually manifest as intensity discontinuities in the 2-D images. Those edges and boundaries represent structures in the image, that are common for multiple image types and can be used for multiple sensor image registration. Edges are defined as points where the modulus of gradient is a maximum in the gradient direction. Along an edge the image intensity can be singular in one direction while varying smoothly in the perpendicular direction. Edges can be created by occlusions, shadows, sharp changes of surface orientation, changes in reflectance properties, or illumination. In IR images of a 3-D scene, most edges represent occlusions and depth discontinuities between objects in the scene, which represent structural information in the image.

5.1 IR image edge detection

A particular difficulty arises in the edge detection for IR/visible image registration. Image registration requires to extract common features which are static in the scene background. In most cases, the background objects in the IR images have the same thermal equilibrium temperature, so that the contrast in the IR image background is related to only the differences in the emissivities and reflectivities of the object surfaces and are therefore very low. Also, the IR images are typically noisy.

Heath *et al*. recently surveyed edge-detection algorithms and assessed their performance[6]. They found that among the five edge detectors: Canny, Nalwa, Rothwell, Bergholm and Iverson, the Canny edge detector usually shows the best performance in the condition that the parameters of the edge detector can be adaptively chosen.

The Canny edge detector is a numerically optimized filter[7]. The optimal filter for step edges can be approximated by the first derivative of Gaussian, which is usually called Canny edge detector. After the filtering, there is a non-maximum suppression process that keeps only the pixels where the values of the output are the local maximum in the direction of the gradient. The values at the neighboring pixels are determined by the linear interpolation. The third process in the Canny detector is the edge linking, which uses a hysteresis thresholding. We first determine edge pixels, which are above a high threshold. Among all other local maxima, which are above a low threshold, we keep only those pixels that are located in the neighborhood of the edge pixels.

The parameters in the Canny edge detector are the width of first derivative of Gaussian filter σ and the low and high threshold values. One problem of the Canny edge detector is its sensitivity to threshold. When the response of an edge point is close to the detection threshold, a small change in edge strength or in the pixellation may cause a large change in edge topology, that makes the extracted edges suspicious, non-reliable, especially near the corners.

The sensitivity to noise is another important problem in the edge detection. The noise in IR images occur as local fluctuations of the image brightness function, which have strong derivative magnitudes, but represent unnecessary image details which are unrelated to image structure. In the IR image background of low contrast with the contrast varying cross the image, the effect of noise becomes important, so that the structural edges may be disrupted and even completely disappear in the edge maps, if a thresholding on the gradient magnitude is applied. The non-maximum suppression in the Canny detector is excessively reliant on the estimation of the gradient angle and so often fails to mark edge pixels at junctions, corners and even on some smooth curve portions where the contrast changes are too poorly defined[8]. This is the reason for broken edges.

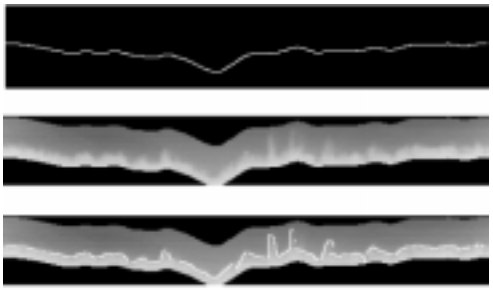
For detecting structural edges in the IR image background, we use the Canny edge detector without thresholding on the gradient magnitude. We avoid the use of threshold on the gradient magnitude, since the contrast is a poor indicator for significance. When the strength threshold is used, of the edges with response close to the threshold, a small change in edge strength or location can cause a large change in the edge topology. We use the large Canny filter of $\sigma \geq$ 6-7, which corresponds to a filter size of 37 - 43 pixels, to obtain the structural edge as a continuous curve, which is the horizon in the scene of battle field, so that the curve length thresholding can be applied to extract the horizon from noisy edges in the edge map. With a small σ, the extracted horizon line is broken. However, with a large σ, the extracted horizon line does not follow the real contour at high curvature. The larger the filter support σ, the less broken the edges are, and, however, the

more image details are filtered out by the large size filter, resulting in a loss of edge localization. Therefore, the multi-scale edge detection is used to recover the localization in the coarse edges.

5.2 Hierarchical Edge Detection

First, the horizon curve is detected at a coarse level with a large Canny edge detector which smoothes the images with a Gaussian of large support σ_0. The horizon is usually the longest curve in the image. For favoring continuity of the extracted curve, no thresholding on the gradient magnitude is applied, such that the horizon appears as a continuous curve or, at least, less broken. Then, the horizon is extracted from the noisy edge map by a curve length thresholding. In the cases where the horizon curves are still broken, we apply the edge saliency measure and combine both edge and region information in order to ensure the extraction of the horizon at the coarsest level, as explained in Section 6.

The coarse horizon is used to guide the search of edges at fine scale. We define a sub-image in the neighborhood of the coarse edge in the original image. The sub-image covers the region along the horizon with 40 pixels above and 10 pixels below each coarse horizon point. The choice of the sub-image size is according to the observation that the images of trees on the hill were cut by the smoothing at the coarse scale. To recover the top of trees we need a search in a large region above the horizon curve. We then apply the Canny edge detector with a small filter width σ within the sub-image. In the experiment, the fine Canny filter was with $\sigma = 0.7$ for visible and $\sigma = 1.5$ for IR images. The noise still exists after the Canny edge detection at the fine scale. However, this noise is within the sub-image zone and may be removed easily by a curve length thresholding, that results in a clearly defined horizon curve. A specific modification on the Canny edge detector was made to prevent the artificially defined sub-image boundaries from appearing as new edges.



(a)

(b)

(c)

Fig.3 Results of Hierarchical Edge Detection

The coarse horizon extracted from an IR image is shown in Fig.3a, where $\sigma_0 = 7.0$, the minimum length threshold applied was 500 pixels. The sub-image is shown in Fig.3b. Figure 3c shows the fine edges obtained by applying a fine Canny edge detector with $\sigma = 1.5$.

The hierarchical edge detection is quit reliable and fast. Since at the fine scale the edge detection is guided by the coarse level edge, the search in large area is avoided, that reduces the computational cost. The shortcoming of the algorithm is the ad-hoc determination of sub-images.

5.3 Edge Focusing

Edge focusing is a coarse-to-fine edge tracking algorithm for recovering the edge points at the finest scale. The scale-space tracking is implemented in a continuous manner. With continuous scaling, the edges are gradually focused by varying the resolution continuously, and moving in the scale space with sufficiently small steps, such that the edge element do not jump farther away than one pixel between successive steps. Our implementation of edge focusing is as following:

1. Detect edge using Canny Detector with the Gaussian smoothing σ_0 sufficiently large so that horizon curve is detected;

2. Extract the horizon using a threshold on the curve length; The horizon curve is denoted as $E(i, j, \sigma_0)$. If (i, j) is an edge point, then $E(i, j, \sigma) = 1$.
3. Detect edges $E(i, j, \sigma_k)$ in a window centered at each edge point $E(i, j, \sigma_{k-1})$, using the Canny edge detector of size $\sigma_k = \sigma_{k-1} - \Delta\sigma$ with $k = 1,2,3...$ and $\Delta\sigma = 0.5$. The window size is 7×7, when $\sigma_k > 2.0$, and is 5×5 when $1.0 \le \sigma_k \le 2.0$, and is 3×3 when $\sigma_k < 1.0$.
4. Go on step 3) until a weak Gaussian smoothing of size σ_K.

In the successive Canny edge detection, after application of the first derivative of Gaussian filter the non-maximum suppression process is applied which keeps only the local maximum in the gradient direction. There is no threshold at finer resolution. The only threshold is on the curve length applied at the coarsest scale σ_0.

Bergholm[9] investigated the deformation of four elementary contour structures: step edge, corner, double edges and edge box. During the edge detection, those contours are generally deformed in four ways: rounding-off, expansion, transformation into circles, or merger, owing to the large Gaussian average operator which blurs the image. In each of the four cases, Bergholm showed that the displacement vector, describing the deformation of the edge contour, is normally of length within the range from 0 to $2|\Delta\sigma|$, where σ is the width of the Canny edge detector, $\Delta\sigma$ is the increment of size of the successive Canny filters. Therefore, if $|\Delta\sigma|$ = 0.5, the displacement of the edge points would be normally less than one pixels, so that corners and junctions may be recovered with a precision less than one pixel.

In real world images we have mostly ramp edges instead of ideal step edges. It is easy to show that the Gaussian blurring operating on a ramp edge always yields smaller displacement than that yielded on a step edge as affirmed by Bergholm. A ramp edge may be modeled as a step edge smoothed by a Gaussian *G* whose size σ_1 depends on the imaging condition and on the camera. Let $r(x, y)$ denote the step edge and $f(x, y)$ the ramp edge in gray level image, then

$$f(x, y) = r(x, y) \otimes G(\sigma_1)$$

where $\otimes$ denotes the convolution. When we use Canny Edge Detector, the image is blurred again with a Gaussian smoothing whose size σ_2 depends on the scale of the edge detector. Let $g(x, y)$ denote the blurred ramp edge before computing the first derivative, then

$$g(x, y) = f(x, y) \otimes G(\sigma_2)$$

therefore

$$g(x, y) = r(x, y) \otimes G(\sigma_1) \otimes G(\sigma_2) = r(x, y) \otimes (G(\sigma_1) \otimes G(\sigma_2))$$

which is equal to

$$g(x, y) = r(x, y) \otimes G(\sqrt{\sigma_1^2 + \sigma_2^2})$$

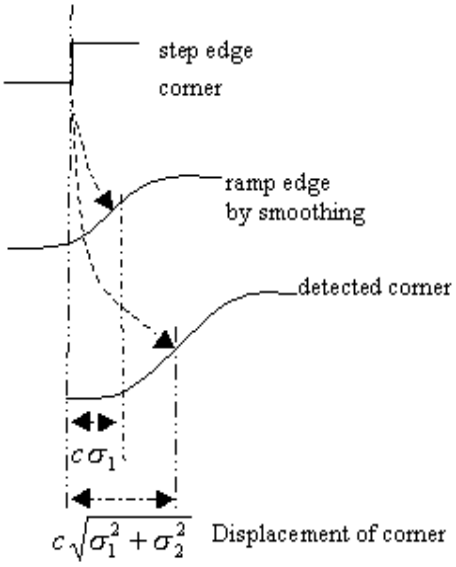
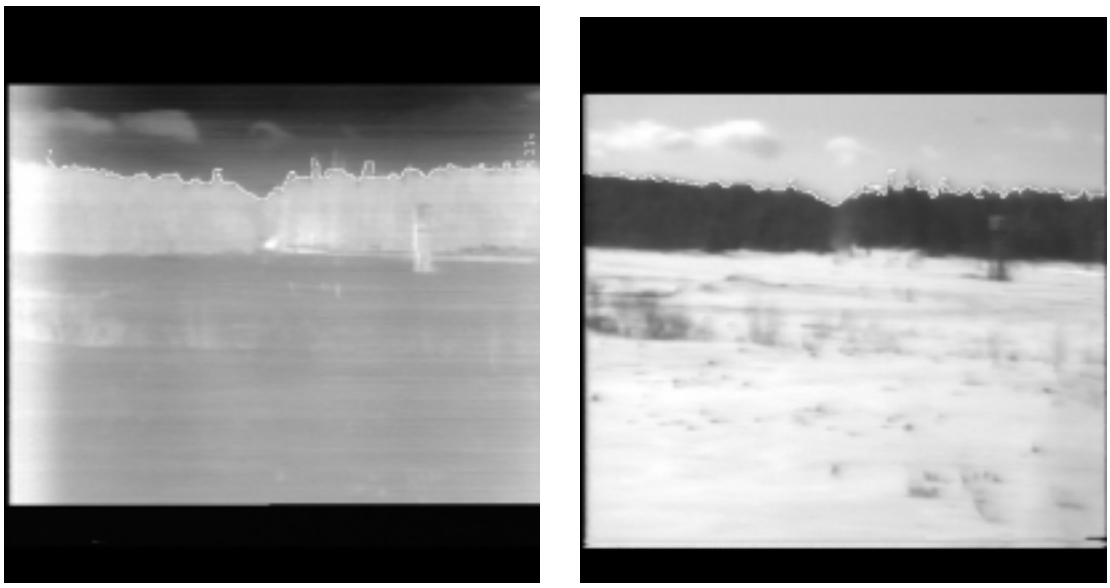


Fig.4 Rounding-off displacement for a ramp edge

Therefore, the length of rounding-off displacement ρ from the corner of ideal step edges to the detected corner is equal to $c\sqrt{\sigma_1^2+\sigma_2^2}$, where c is a constant. However, the displacement from the center of the ramp corner to the detected corner would be proportional to $\sqrt{\sigma_1^2+\sigma_2^2}$ - σ_1, as illustrated in Fig. 4 and would be less than σ_2. Therefore, if $|\Delta\sigma_2|$ = 0.5 in the edge focusing, the displacement of the ramp edge corner would be less than one pixels.

In our IR images the ramp edges of trees can be very slow of more than 20 pixels wide, corresponding to a large σ_1 more than 10. The edges around the trees were cut completely when a Canny edge detector of σ_2 = 7 was applied. This is because the large displacement of the corner $\sqrt{\sigma_1^2+\sigma_2^2}$. However, using the edge focusing we were able to recover the edges and tops of the trees, which would be important for the image registration.

We implement the edge focusing algorithm with the filter size increment $\Delta\sigma = 0.5$ and varying size windows. We chose to use the window size larger than the usually used, 3 x 3, so that the gradient magnitude values can be evaluated at the two neighboring pixels, because in the non-maximum suppression the determination of an edge pixel requires to compare with at least two neighboring pixels. We believe that the length of rounding-off displacement ρ can be larger than one pixel, because the real ramp edges in our IR images were noisy and do not follow the theoretical model described in the precedent.



(a) (b)

Fig.5. Experiment results of Edge Focusing. (a) Visible image. (b) Infrared image.

For images shown in Fig.5, we first detected the coarse horizon with $\sigma_0 = 4.5$ for visible image and $\sigma_0 = 7.0$ for IR image using Canny Edge Detector. Then we applied the edge focusing with the scale step $\Delta\sigma = 0.5$ and the varying size windows. The final scale was $\sigma = 0.7$ for visible image and $\sigma = 1.5$ for IR image. Figure 5 shows the extracted edges which follow nicely the silhouette of the hill with some flat tops of trees recovered in both visible image and infrared image.

6. DETECTING HORIZON WITH CURVE SALIENCY MEASURE

The Edge detector is basically a local operator. However, the structural edges useful for image registration are not local features, but exhibit regional and global nature in many cases. Salient structures can often be perceived in an image at a glance[10]. They appear to attract our attention. Therefore, we use curve saliency measure to help detecting the structural edges.

When cameras are mounted on a grounded vehicle, it is reasonable to assume that the camera axis is pointing approximately horizontally. Unless the terrain is very steep (or the vehicle is driving alongside a wall) the horizon is usually visible in the image and is usually the most distant part of the scene (except for the sky). The horizon lines are common in the IR and visible images, and are independent of the grayscale level disparities and contrast polarity reversals. More importantly, it may be possible to segment the horizon line from noisy edges on the ground by a threshold of curve lengths, since the horizon line has the longest length in the image. The fact that the horizon is the most distant part in the image helps fitting the distortion transformation for image registration.

Salience measures can be region-based or curve-based. In the visible images the horizon bounds the brightest part of the image, which is usually always the sky. Duric and Rosenfeld[11] use the horizon detection for stabilization of image sequence from a ground vehicle. They detected the horizon by finding the bright parts of the image (sky) and then estimating the boundaries of these parts. This approach uses then the regional information. We attempt to use the curve-based salience measure to detect horizon lines in IR and visible images. The curve saliency measure is defined to favour long over short curves and smooth over wiggly curves. For horizon, we define the salience measure which is estimated at each pixel along a curve *i*, as

$$\Phi_i = \frac{L_i}{N} \sum_{k=1}^{N} \sigma_k \frac{|y_{k+1} - y_k|}{|x_{k+1} - x_k|} + \alpha_k$$

where N is the total number of pixels on a segment of the curve *i*, whose horizontal extension in x is L_i

$$\alpha_k = \begin{cases} 1 & if \quad y_{k+1} - y_k = 0 \\ 2 & if \quad y_{k+1} - y_k \neq 0 \end{cases} \qquad \sigma_k = \begin{cases} 0 & if \quad x_{k+1} - x_k = 0 \\ 1 & if \quad x_{k+1} - x_k \neq 0 \end{cases}$$

The horizon in the natural scene usually is not a horizontally straight line. This curve saliency measure favorites inclined segments rather than horizontal or vertical ones. The horizontal segments have contribution of 1 to the saliency Φ_i, vertical segments have saliency of 2, since $\sigma_k = 0$. Inclined segments have saliency of 3. However, wiggly curves will receive small total saliency measure because of small L_i. We evaluate the salience measure for each curve in the edge map and retain 3 - 4 most salient curves. Then, we fill gaps between the salient curves and re-evaluate the saliency of the connected curves. With this approach we can detect the horizon at a coarse scale without thresholding of curve length, so that the horizon is detected even it is broken by noise into several segments and contains a number of gaps.

7. IMAGE ALIGNMENT USING HAUSDORFF DISTANCE

The purpose of image alignment is to register a pair of images such that the extracted static scene features are optimally aligned. Feature based image registration requires a specification for the features, the parameter space, the image transformation, which aligns the image, and the search strategy for finding the best alignment according to some objective function. For aligning images from different modalities, edges arising from depth discontinuities can be considered as most salient. Given a set of salient edges from each image, the next step is to determine the image transformation which aligns those features considered to be a static reference for the scene. The search for the optimal image transformation can be implemented in several ways. The methods can be classified as feature matching methods or transformation space methods.

Feature matching methods determine the correspondence between the elements of the feature sets, i.e., corresponding features are projectively given by the same scene feature. The transformation space methods search the parameter space for the solution that achieves an optimal alignment of the static projected scene features. The drawback of feature matching methods is the prohibitive cost of detecting and eliminating outliers, i.e., features which do not have a match. Its advantage is that once a set of correct matches is found the image transformation is, in general, quick to compute. Transformation space methods can be prohibitively expensive because the search space is generally very large, however, outliers are easily handled by using rank order statistics. A strategy for efficiently searching the parameter space is given by Huttenlocher *et al.*[12] In view of the large proportion of outliers in feature based multi-modal image alignment a transformation space method based on the directed Hausdorff distance was implemented. The size of the search space is reduced by partitioning the image into blocks and searching for translations that minimize the Hausdorff distance between corresponding blocks. The assumptions are that the motion can be locally approximated by simple translations of blocks, and the percentage of outliers and an error bound for the feature alignment are known approximately.

The Hausdorff distance is defined by

$$A = \{a_1, \ldots, a_m\} \text{ and } B = \{b_1, \ldots, b_n\}$$
$$H(A, B) = \max(h(A, B), h(B, A))$$
$$h(A, B) = \max_{a \in A} \min_{b \in b} \|a - b\|$$

where A and B are point sets, H is the generalized Hausdorff distance and h is the directed Hausdorff distance. In the presence of outliers the Hausdorff distance will return the greatest distance which is likely due to an outlier. To be able to compare portions of the data sets the partial directed Hausdorff distance is defined,

$$h_k(A, B) = \underset{\mathrm{a} \in \mathrm{A}}{k\text{th}} \min_{b \in B} \|a - b\|.$$

This expression evaluates to the k^{th} ranked distance.

The alignment method using Hausdorff distances proceeds as follows for a pair of images after extraction of the salient edges

1) Compute a quadtree partition of each edge image such that no block without edge points is further subdivided. The partition with fewer blocks is retained for both images. Define a set of model edge points for the first block in image 1 from the edge points that lie within that block. Create a model image from these edge points.
2) Define a set of subimage edge points from the corresponding block in image 2 from the edge points within the block extended by a border whose dimensions correspond to the largest expected vertical and horizontal displacements. Create a target image from these edge points.
3) Compute the directed partial Hausdorff distance under a translation transformation from the model image to the target image. The translation which minimizes the kth ranked distance is retained. The search strategy in the translation parameter space is described in Huttenlocher et al.12
4) Repeat steps 2 and 3 for the remaining non-empty blocks. If at least 3 blocks provide local translation estimates from step 3 then the global affine transformation is estimated, the nonreference image is resampled according the global affine transformation and the images are fused. The image fusion is accomplished by an appropriately weighted combination of the aligned images brightness values.

Fig. 6 shows a scene taken simultaneously by a daylight and IR camera at Defense Research Establishment Valcartier. The viewpoints of the two cameras are displaced slightly and there is a slight relative rotation about the optical axis which would yield a very poor fused image if no alignment is made. The quadtree decomposition stops at the first level, i.e., there are 4 blocks. The salient edges include the silhouette of the hill and some ground structure, which are overlayed on the images.

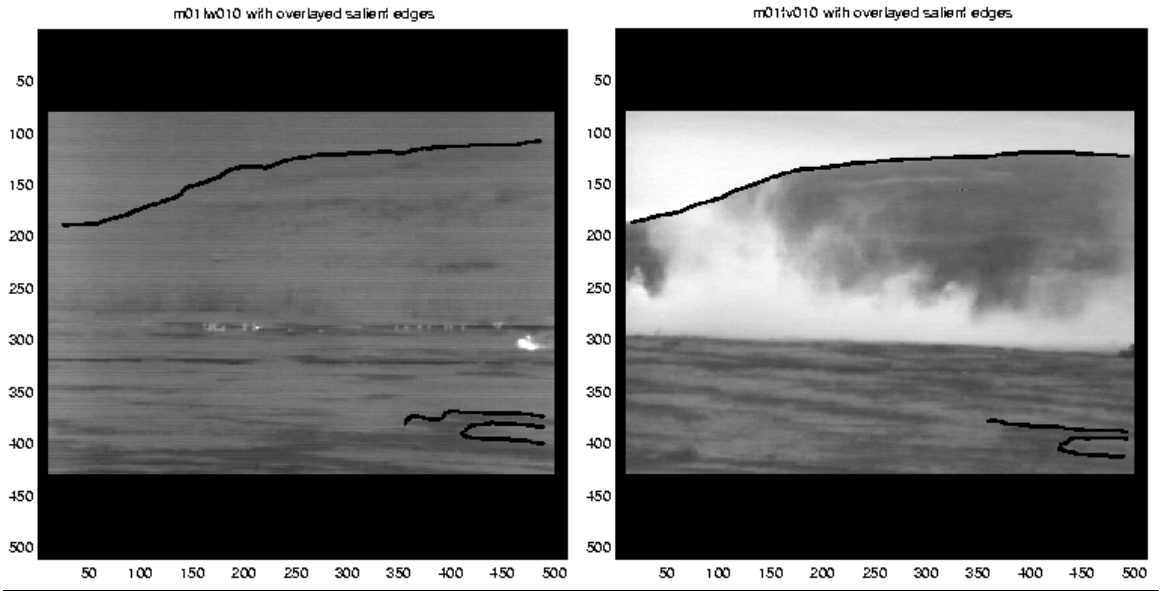


Figure 6. Infra-red and visible images of a scene with soldiers hidden by smoke and a truck.
Salient edges for image alignment are overlayed.

Finally, Fig. 7 shows the fused aligned images. The salient edges are registered in each of 3 blocks, the fourth block contains no edge points. The Hausdorff distance is used to find the optimal displacement assuming 5 percent outliers for the blocks covering the hill edge and 10 percent outliers for the edge in the lower right block. The specified search strategy finds the translation for each block such that 90 percent of the visible image edge points are no more than 5 pixels from some IR image edge point for the corresponding block. The local displacements are then used to determine the global affine transformation to register the two images. The estimated (x,y) displacements for the blocks upper left, upper right and lower right that are supplied to the global affine estimator for aligning the visible image to the IR image are (33,-3), (-11,-7) and (-8,-11) respectively.

Figure 7. Aligned and fused IR and visible images.
Fusion is by weighted combination of image brightness values after alignment

The estimated affine transformation parameters that map point p in the visible image to the point p'in the IR image such that $\mathbf{p'} = \mathbf{Mp}+\mathbf{t}$ are

$\mathbf{M} = \begin{bmatrix} 0.8239 & 0.0544 \\ -0.0179 & 0.9897 \end{bmatrix}$ and $\mathbf{t} = (17.1925, -6.2471)^T$.

Note that the image coordinate system origin is top left with positive x to the right and positive y down.

8. CONCLUSION

We have analyzed the problems in the real world visible/IR image registration. The area-based approaches are still feasible. However, feature extraction of structural edges as common features for registration and feature matching methods are more powerful to process with our low quality IR images. We have implemented multi-scale hierarchical edges detection and edge focusing and introduced a new salience measure for the horizon. For multisensor image registration, the common features extracted from images of two modalities can be still different in detail. Therefore, the transformation space match methods with the Hausdorff distance measures are more suitable than the direct feature matching methods. We have introduced image quadtree partition technique to the Hausdorff distance matching, that dramatically reduces the size of the search space into that of the search for translations which minimize the Hausdorff distance between corresponding blocks. We have shown image registration of visible/IR real world images of battle fields. The key point is to extract salient features from the real world images using local, regional and global information and appropriate salience measures.

---

[1] Burt P. J. and Adelson E. H., "The Laplacian pyramid as a compact image code", IEEE Trans. on commun. Vol. com-31, No.4, 532-540 (1983).

[2] Sharma R. K. and Pavel M., "Registration of video sequences from multiple sensors", Proc. Image Registration Workshop, NASA Goddard Cenetr, 361-364 (1997)

[3] Brown L. G., "A survey of image registration techniques". ACM comput. surveys, vol.24, 325-376 (1992).

[4] L. Sevigny, "RSG.9(panel3) Canadian test sequences", DREV internal report (1997).

[5] Irani M. and Anandan P., "Robust multi-sensor image alignement", Image Understanding Workshop, 639-647 (1997).

[6] Heath M. D. et al., "A robust visual method for assessing the relative performance of edge-detection algorithms", IEEE Trans. PAMI vol.19, No.12, 1338-1359 (1997).

[7] Canny J., "A computational approach to edge detection", Trans. IEEE PAMI-8, No.6, 679-698 (1986).

[8] Rothwell C., Mundy J., Hoffman B. and Nguyen V., "Driving vision by topology", Report No. 2444, INRIA, France (1994).

[9] Bergholm F., "Edge focusing", IEEE Trans. PAMI-9, No.6, 726-741 (1987).

[10] Ullman S. and Sha'ashua A., "Structural saliency: the detection of Globally salient structures using a locally connected network", A.I. Memo No. 1061 MIT (1988).

[11] Duric Z. and Rosenfeld A. "Image sequence stabilization in real time", Real-Time Imaging, vol. 2, 271-284 (1996).

[12] Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., "Comparing images using the Hausdorff distance," IEEE Trans. PAMI-15, No.9, 850-863 (1993).