Stabilization of Infra-red Aerial Image Sequences Using Robust Estimation

Daniel McReynolds[a], Pascal Marchand[a], Yunlong Sheng[a]
Langis Gagnon[b], Léandre Sévigny[c]

[a]COPL, Dept. de Physique, Université Laval, Ste-Foy, Qc, G1K 7P4, Canada
dmcreyn,sheng@phy.ulaval.ca   marcha00@gel.ulaval.ca

[b]Lockheed Martin Electronic Systems Canada, 6111 Royalmount Ave., Montreal, Qc, H4P 1K6, Canada[δ]
langis.gagnon@crim.ca

[c]Defence Research Establishment Valcartier, 2459 Boul. Pie XI nord, Val-Bélair, Qc, G3J 1X5, Canada
leandre.sevigny@drev.dnd.ca

Abstract

Image stabilization is the image registration applied to one video image sequence from a single camera, which has been identified as the key first step for the task of multi-spectral image fusion for aerial surveillance applications. The stabilization of infra-red (IR) aerial image sequences is challenging owing to the low contrast and signal-to-noise ratio of the images, and to potentially large viewpoint changes that result in images with large rotation and scale change and perspective distortion. We demonstrate a new feature-based method for IR aerial image sequence stabilization. We use the greylevel differential invariant (GDI) matching due to Schmid and Mohr which is invariant to rotation and scaling. Extensions to the basic GDI method are introduced that improve the performance of the method. We use M-estimation for the image registration parameters. The method is robust to outliers returned by the GDI method. We verify the point correspondence under orthographic projection using the epipolar constraint. Experimental results are reported for real-world and synthesized IR image sequences.

1 INTRODUCTION

Imaging sensors of different modalities, for example, visible and infra-red, provide complementary information which, if properly fused, can assist the observer in the image interpretation task. Image stabilization is a key step in this process. Image stabilization is the image registration applied to one video image sequence from a single camera. When the camera is mounted on an unsteady or a moving platform and objects are far from the camera, the 3-D space motion of the camera will affect the images.

Until now, the primary means available to stabilize images from a camera on a moving vehicle has been to mount the camera on an electro-mechanical stabilizing platform. These stabilizers are bulky and expensive. Their performance degrades with vibration in the critical $0 - 20$ Hz range. An alternative stabilization method uses image processing techniques to first estimate and then eliminate the scene motion that is due to camera motion by warping each image frame into precise alignment with a reference frame.

The image stabilization task has been specified in general terms as follows. Given an image sequence usually consisting of at least 10 to 30 seconds of data, a special frame called the reference frame is chosen at or near the beginning of the sequence. Frames subsequent to the reference frame shall be registered to the reference frame in such a way that the frames are precisely aligned with the reference image.

The conventional image registration method uses block matching techniques. The block matching utilizes the full image information and can be applied to any type of image, rich or poor in texture. The block correlation is robust against random noise and has high accuracy. However, block matching can account for only translations and only approximately for other image distortions. It is expensive to compute and becomes prohibitive when the image displacement is large. Also, the cross-correlation based on image intensity similarity is sensitive to environment changes and is not applicable for registration of multi-spectral images [2].

On the other hand, feature-based image matching can account for any image deformation and can be insensitive to multi-sensor modalities by selecting structurally salient features. It is quick to compute. However, the feature-based methods will fail to find matches in structure-less areas. Its reliability depends on the feature extraction

---

process. Therefore, robustness of the matching process is a critical issue.

The only known geometric constraint between two images of a single scene is the epipolar constraint. The epipolar geometry describes relations between corresponding points in two images, and is, therefore, useful for image matching. Zhang *et al.* [18] proposed a robust technique for image matching for uncalibrated cameras based on geometric verification using the epipolar constraint. They use cross-correlation between local supports of the feature points for the initial matching, which is not invariant to scaling, rotation and other deformations, and is the main limitation of their approach. They use Least Median of Squares (LMedS) to discard outliers, that allows up to 50% false initial matches in the estimation of the epipolar geometry. Our experiments show that Zhang's approach can provide a large number of correct matches for outdoor images from a ground based camera where the perspective projection results in local translations of the corresponding points. However, Zhang's approach fails in its initial matching with aerial images where there is, for example, a large global rotation of the input scene.

In this paper we propose an approach for the stabilization of real-world aerial infra-red image sequences. In the aerial image sequences the frame to frame motion can be very large. The IR images are generally noisy with low contrast. We view the task of image stabilization as a problem in image registration, that can be accomplished using image matching methods. A classic approach to image matching is to hypothesize then verify the correspondence of features between images [9]. In this approach, the initial matching is critical for a successful matching and registration. To improve the initial matching by cross-correlation Deriche [4] added the directions of the gradient, curvature, and a disparity measure. Hu [7] suggested to apply the cross-correlation in several directions and Ballard [1] used steerable filters. Schmid and Mohr [12] developed a point matching method based on greylevel differential invariants (GDI), which is invariant to image rotation, scaling and translation. We extract feature points using the Harris-Stephens corner detector [6], which is efficient to compute and has been shown to be one of the best detectors in terms of repeatability with viewpoint change and other scene variables [14]. It is efficient also to extract corner points in the texture of the IR images. We use the greylevel differential invariant matching method to find an initial set of matches. Three additional enhancements to the basic greylevel differential invariant matching method are introduced and tested: (1) Search over the GDI space with *k*-d trees to speed up the matching process, since a query finishes in logarithmic expected time, (2) Scale-space verification of matches to improve the ratio of true to false matches, and (3) Determining the covariance matrix for normalizing differential invariants at runtime.

The GDI method provides a set of matches that often contain outliers which can seriously perturb the estimation of the image transformation. We use M-estimation which is robust to outliers and is computationally efficient, and a variant of least median of squares, which is more computationally expensive, in the case that M-estimation failure is detected. For aerial images, the distance from scene to camera is much larger than the depth of the scene, and the field of view may be small. In this case, the orthographic projection provides a good approximation to perspective projection, as shown by Pritt [11] as well as Shapiro *et al.* [15], and then the unique optimal solution, in a least squares sense, for registering a pair of images, is the global affine transformation with local distortion modeled by a displacement along parallel epipolar lines proportional to the height of the scene point [11]. However, Pritt does not address the critical issue of the initial matching, estimation of the transformation parameters in the presence of outliers, nor are results presented for real world image data. We fit the GDI matches to the global affine transformation, that accounts for image scaling and rotation in the important initial matching stage and implement a robust image registration. It is by incorporating robust estimation into the image registration process that we perform a match verification process using the epipolar constraint.

2 Enhanced greylevel differential invariant matching

At each feature point in the image found by the Harris-Stephens corner detector, a greylevel differential invariant (GDI) vector is computed. The GDI representation and matching method are invariant to image rotation, scaling and translation [12]. The representation is fairly robust with respect to rotation in depth that leads to foreshortening of surface patches, i.e., in general, a local affine distortion of the brightness surface. The greylevel differential invariants are based on the derivatives of the Gaussian filtered image in a representation called the local jet. The local jet represents the truncated Taylor series expansion of the image function and is useful for encoding the local geometry of the brightness surface. The use of Gaussian smoothing makes the differentiation well posed in addition to having other nice mathematical properties.

The details of the greylevel differential invariant representation can be found in [12]. The components of the GDI vector representation are rotationally symmetric functions. Differential invariants can be made invariant to an affine transformation of the brightness function by an appropriate normalization that yields a vector of 7 components.

To give scale invariance over a fixed range we compute a set of scales centered on a reference scale, σ_0, such that

$$\sigma_i = (6/5)^i \sigma_0$$ where i $\in$ (-*n*... -1, 0, 1... *n*). A value

of four for n yields the scale factor range 0.48 to 2.07, hence there are nine differential invariant vectors for each keypoint. This is a multi-scale representation. The 20 percent factor is empirically derived and reflects the expected differential scale range over which the invariants do not change appreciably.

Nearest neighbour search by k-d tree The Mahalanobis distance is used to determine the nearest neighbour. Points are declared to be corresponding when the pair of points from two images are mutually selected as closest. The space of differential invariant vectors can be organized in a hash table [12] or with a tree representation such as the k-d tree. Nearest neighbour searching with k-d trees reduces the search time to logarithmic expected time. The expected number of records examined to satisfy a query is given by the expression $R(k,b) \approx b\{[G(k)/b]^{1/k} + 1\}^k$, where k is the record dimension, b is the number of records in a bucket, and $G(k) \geq 1$ is a constant that accounts for the geometric properties of the norm used for the distance measure [16]. For our implementation of GDI matching, $G(k)=1$, k is 7 and b is 1, hence $R(7,1)=128$.

Match verification in scale-space Experimental results suggest that nearest neighbour matching with differential invariants is scale sensitive over moderate viewpoint changes. A local multiscale analysis is used to filter out scale-space unstable and hence potentially incorrect matches. The full differential invariant matching process is carried out at three reference scales that differ by multiples of ten percent of the reference scale σ_0, i.e., $\sigma_0(1 + k * 0.1)$ for $k=0,..,2$. The value of ten percent is half the expected scale sensitivity of differential invariants. This value was found experimentally to yield good results. Too large a value eliminates most matches and too small a value does not effectively eliminate scale unstable matches. A match is scale-space verified if it exists at all three scales. Experimental results demonstrate that this verification step significantly increases the ratio of correct to incorrect matches. This step increases the computational cost by a factor of three but it is parallelizable.

Normalization The variant of k-d tree used here requires that the data satisfy a Euclidean norm. An appropriate transformation of the GDI vectors by a suitable covariance matrix accomplishes this. The covariance matrix describes the expected variance of the differential vector components with respect to changes in viewpoint, illumination, sensor properties, and noise. For an image retrieval task [13] the covariance matrix is determined by averaging together over all keypoints the covariance matrix generated for each keypoint by tracking observations over an extended image sequence. It is not clear, however, if this is suitable for the image matching problem. First, we assume that tracking is generally not an option. Secondly, the resulting covariance matrix may not be suitable for all image pairs to be matched, except, of course, the images from the sequence itself. Computing and combining covariance matrices over several different sequences may be sufficient for a variety of image retrieval tasks.

Instead, we adopt a classical perturbation approach to estimating the covariance matrix for the differential invariant vectors. Such an approach is frequently used when an analytic solution is not feasible. For example, this approach is used to determine datum/model compatibility for the RANSAC algorithm [5]. The RANSAC procedure uses data perturbation to estimate implied error bounds for a given model. We compute a covariance matrix directly from the given image pair by modeling brightness variation due to all the factors mentioned above by normally distributed noise added to the images. The noise variance is some fraction of the variance of the image brightness over the entire field of view. A value of 25 percent of the brightness variance has been used in the described experiments. This is an empirical value that has yielded good results. Furthermore, the matching results are not overly sensitive to this value. The final covariance matrix is then estimated by averaging together a large number of covariance matrices generated from differential invariant vectors computed from images with and without added noise.

3 Image registration with M-estimators

The following steps outline the registration algorithm.

1. Detect Harris-Stephens corners with derivative of Gaussian filters, $\sigma = 1$ pixel, for reference and non-reference images.
2. Low pass filter reference and non-reference images $\sigma = 4$ pixels.
3. Perform GDI matching on detected corners yielding GDI matches. Baseline filter scale is 5 pixels. If < 4 matches then FAILURE and go to step 9.
4. Eliminate gross outlier matches. The gradient angle test deletes those matches that have an angle change of a magnitude greater than $45°$ from the median angle change for all matches. These matches are defined as the gross outliers. Remaining matches are called candidate matches. If < 4 candidate matches then FAILURE and go to step 9.
5. Estimate p_1 to p_6 with Fair weight function. If iterations do not converge then FAILURE and go to step 8 (or return to 8c if previous step was 8b).
6. Initialize Tukey bi-weight function weights using parameter estimate from previous step.
7. Estimate p_1 to p_6 with Tukey bi-weight function. If iterations do not converge then FAILURE.
8. If FAILURE or $\geq 50\%$ of weights $= 0$ during Tukey iterations then
   8a) Select at random 4 matches.
   8b) Execute steps 5 to 7.
   8c) If any weight $= 0$ or FAILURE then discard result else store parameter estimate and median residual over candidate match set.

8d) Repeat 8a-c until 71 subsets found or until iteration limit exceeded.

8e) If 71 subsets found then do steps 6 & 7 with lowest median residual parameter estimate else FAILURE.

9. If not FAILURE then resample non-reference image as registered frame according to parameter estimate else copy reference frame as registered frame.

10. If no more frames then QUIT else return to step 1 processing next non-reference image only.

3.1 Epipolar Constraint

The only known geometric constraint that exists between corresponding point matches from two or more views is the epipolar constraint, which can be used for image matching. When the field of view is small and variation in depth of the scene is small compared to its average distance to the camera, the orthographic and scaled orthographic projection is a close approximation to the perspective projection model [11][15]. The approximation is also good for large field of view if the separation angle between two cameras is small [11]. Under this approximation, the model locates the optical center at infinity, hence, the projection rays are parallel to one another and perpendicular to the image plane. The epipoles are situated at infinity in the image plane and the epipolar lines are parallel.

The parallel projection is linear with the registration function expressed by [11]

$$\mathbf{F}(x_1, y_1) = \mathbf{A}(x_1, y_1) + h(x_1, y_1)\mathbf{e} \qquad (1)$$

where $\mathbf{F} = (x_2, y_2)$ maps points (x_1, y_1) from the first image to the second, $\mathbf{A}(x_1,y_1)$ is an affine transformation given by

$$\begin{pmatrix} \hat{x}_2 \\ \hat{y}_2 \end{pmatrix} = \begin{pmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \qquad (2)$$

and $h(x_1, y_1)$ is the z coordinate of the non-occluded point in the input scene. Its value depends on the choice of the scene coordinate system. The vector $\mathbf{e}$ is the epipolar vector. The parallel epipolar lines are expressed as $\mathbf{A}(x_1,y_1) + \alpha\mathbf{e}$ which depends only on the position and orientation of the two image coordinate systems relative to the scene coordinate system and α is a real number. The registration function is globally defined by an affine transformation. Eq. (1) accounts explicitly for the depth variation via the height function $h(x_1,y_1)$ as a local distortion which are displacements along epipolar lines and proportional to the heights in the scene.

The choice of parameters $\mathbf{A}(x_1,y_1)$, $h(x_1,y_1)$ and $\mathbf{e}$ in Eq. (1) is not unique. The solution becomes unique with normalization constraints proposed by Pritt, namely that the zero and first order moments of the scene's height function $h(x_1,y_1)$ be equal to zero. Under this constraint, all

the terms involving h disappear in the equations for the least mean square solution of Eq. (1) yielding the least mean square estimate of the affine transformation $\mathbf{A}(x_1,y_1)$. The affine component of the registration function determines the least squares estimate of a planar approximation of the scene structure. After the determination of $\mathbf{A}(x_1,y_1)$, the epipolar vector $\mathbf{e}$ and the scene height function can be recovered from the residuals:

$$s = x'_2 - [p_1 x_1 + p_2 y_1 + p_3] \qquad (3)$$

and

$$t = y'_2 - [p_4 x_1 + p_5 y_1 + p_6] \qquad (4)$$

where (x'_2, y'_2) is the observed point in the second image that corresponds with (x_1,y_1). The slope of the epipolar vector $\mathbf{e}$ is given by

$$m = [\beta + \delta(\beta^2 + 4)^{1/2}] \div 2 \qquad (5)$$

where $\beta = \Sigma(t_i^2 - s_i^2)/ \Sigma(s_i t_i)$ and $\delta = sgn(\Sigma(s_i t_i))$. This estimate for the slope of the epipolar line is the least squares solution for the objective function that minimizes the orthogonal distances of the points (s_i, t_i) to the line. The robust estimation of the affine transformation parameters therefore implements a verification of the matches using the epipolar constraint. The epipolar vector $\mathbf{e}$ is given by

$$\mathbf{e} = [e_1, e_2]^{\mathrm{T}} = [1/(1+m^2)^{1/2}, m/(1+m^2)^{1/2}]^{\mathrm{T}}. \qquad (6)$$

The solution for h_i is given by

$$h_i = s_i e_1 + t_i e_2. \qquad (7)$$

3.2 M-estimation applied to image registration

Let $r(s_i, t_i)$ be the ith residual, i.e., the difference between the ith observation and the current estimate of the fitted value. The classic least squares problem seeks to minimize $\sum_i r_i^2$ which is known to be unstable in the presence of outliers in the data. The M-estimators are designed to reduce this instability by replacing the squared residuals r_i^2 by a function of the residuals which yields the problem [17]

$$\min \sum_i \rho(r_i) \ ,$$

where ρ is a symmetric, positive definite function with a unique minimum at zero and is chosen to be less increasing than square. The problem is not solved directly but is implemented as an iterated re-weighted linear least squares solution.

We seek to estimate p_j for $j=1$ to 6 from Eq. (2). We solve for the parameters p_1 to p_3 using the M-estimation formulation

$$\sum_i w(r_{xi}) r_{xi} \frac{\partial r_{xi}}{\partial p_j} = 0 \qquad (8)$$

over $i=1,\ldots,k$ data points and $j=1,2,3$.
Similarly for p_4 to p_6 we solve

$$\sum_i w(r_{yi}) r_{yi} \frac{\partial r_{yi}}{\partial p_j} = 0 \qquad (9)$$

over $i=1,\ldots,k$ data points and $j=4,5,6$.

3.3 Choice of objective functions

Some ρ-functions assure a unique solution, but since they only reduce the effect of outliers, the estimator can be biased. Other ρ-functions do not guarantee a unique solution, but reduce considerably the effect of outliers or even more, eliminate them. Since no ρ-function is perfect, we use both types of ρ-functions as proposed by Huber [8]. We began the estimation process with the objective function Fair and then refine the estimate with Tukey's bi-weight function. For both functions, the convergence threshold used is 0.1% in the relative difference between iterated parameter estimates. The Fair distribution has continuous derivatives up to third order and yields a unique solution.

Given these choices for the objective function the estimator of the global affine transformation can be written as follows from Eq. (8) for the residual in the x component

$$\sum_i w_i^j [x'_{2i} - (p_1 x_{1i} + p_2 y_{1i} + p_3)](-x_{1i}) = 0$$
$$\sum_i w_i^j [x'_{2i} - (p_1 x_{1i} + p_2 y_{1i} + p_3)](-y_{1i}) = 0 \qquad (10)$$
$$\sum_i w_i^j [x'_{2i} - (p_1 x_{1i} + p_2 y_{1i} + p_3)](-1) = 0$$

where $\dfrac{\partial r_{xi}}{\partial p_1} = -x_1$, $\dfrac{\partial r_{xi}}{\partial p_2} = -y_1$, and $\dfrac{\partial r_{xi}}{\partial p_3} = -1$. The jth

iteration of the weight, w_i^j, is one of Fair or Tukey's bi-weight as described above. Similarly, for the y component from Eq. (9)

$$\sum_i w_i^j [y'_{2i} - (p_4 x_{1i} + p_5 y_{1i} + p_6)](-x_{1i}) = 0$$
$$\sum_i w_i^j [y'_{2i} - (p_4 x_{1i} + p_5 y_{1i} + p_6)](-y_{1i}) = 0 \qquad (11)$$
$$\sum_i w_i^j [y'_{2i} - (p_4 x_{1i} + p_5 y_{1i} + p_6)](-1) = 0$$

where $\dfrac{\partial r_{yi}}{\partial p_4} = -x_1$, $\dfrac{\partial r_{yi}}{\partial p_5} = -y_1$, and $\dfrac{\partial r_{yi}}{\partial p_6} = -1$.

Eqs. (10) and (11) give a linear system of 6 equations in the 6 unknown parameter variables p_1 to p_6 which are solved for iteratively in two phases. The first phase uses the Fair weighting function to yield an initial

estimate. The final estimate uses the Tukey bi-weight function with weights initialized with the parameters from the Fair estimate.

The registration function for the experimental results reported here is given by $\mathbf{F}(x,y) = \mathbf{A}(x,y)$, i.e., the height term is ignored. The assumption is that the height function values are not large enough to significantly perturb the registration for the aerial image sequences. It is straightforward to generalize the approach by the addition of the term $h(x,y) \cdot \mathbf{e}$ to the global affine transformation. The drawback is the necessity to estimate $h(x,y)$ over the entire field of view. This requires us to interpolate the height function from known point correspondences which will yield a poor estimate for a sparse set of point matches or, alternatively, one can compute a dense matching using a method such as optical flow. Match density for point features can be increased by searching for additional matching points along the estimated epipolar lines, a one dimensional search.

3.4 Ill-conditioned solutions

In some rare cases, convergence given by the Fair function is not the one we would expect. This can happen when the GDI method returns matches where the number of outliers is not relatively small or their distribution perturbs the estimator significantly. In this case, the minimum reached depends upon the influence of the outliers and the initial value of the residual function. The Fair function may converge to a minimum such that many of the matches have relatively small weights. After convergence of the Fair function, Tukey's bi-weight function seeks to improve the estimate of the parameters by eliminating the effect of large outliers. If we then look at the $\omega(i)$, we see that the majority of the matches are zero weighted.

To override this situation, we added a least median of squares process which provides a new initial parameter estimate that is then passed to the Tukey's bi-weight function process for computing a final estimate. We defined as a threshold for computing a revised initial estimate that at least 50% of all matches must be 0-weighted from the final estimate provided by M-estimation with Tukey's bi-weight function. The process that finds a better initial estimate randomly chooses 4 matches, which is the minimum number of matches to do a minimally over-constrained estimate, from the set of candidate matches. These four matches then pass through the Fair and Tukey's bi-weight estimation processes as before. The residual for all candidate matches is computed according to the parameter estimate found for this subset. The values of the estimated parameters and the median residual are stored. If one or more of the subset $\omega(i)$ is 0, the subset result is ignored. The process iterates until 71 subsets are selected. This number of subsets guarantees with a

probability of 0.99, assuming no more than 50 percent outliers, that at least one of the subsets is composed of 4 good matches. A final estimate is made involving all the matches using the Tukey's bi-weight function. The $\omega(i)$ are initialized with respect to the Tukey function according to the parameter estimate with the smallest median residual. This gives a more robust estimate of the global affine transformation parameter vector. Note that this random search process is more computationally expensive than the M-estimation process and, therefore, is not used as the main estimation method.

4 Experimental results

4.1 Infra-red aerial image sequence DIM01

Figure 1 gives the GDI matching results for frames 01 and 20 from the sequence DIM01, a sequence provided by Defence Research Establishment Valcartier for algorithm development purposes. The sequence is acquired at a 4Hz frame rate by a helicopter mounted infra-red camera. Note the large scale change due to the camera translation in the foreground and a rotation about a point in the lower left image quadrant. The scale of the smoothing kernel for the differential invariants is five pixels with Gaussian pre-smoothing of the images at a scale of four pixels to reduce the IR scan noise as well as to compensate for the lack of fidelity of the actual image transformation with the assumed transformation. The GDI model assumes the image is transformed by a similarity transformation. The actual model is better approximated by a local affine transformation that in turn is a local
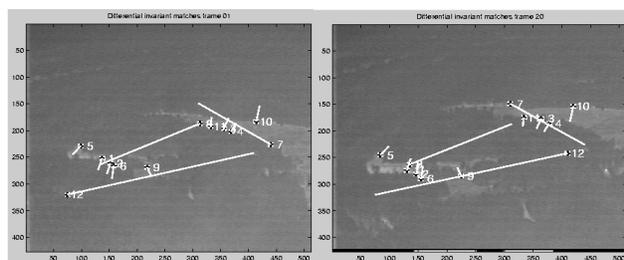
Figure 1: Greylevel differential invariant matches. Left image is frame 01 and right image is frame 20. Matches 7, 8 and 12 are incorrect.

approximation of the effect of the perspective distortion due to the translating camera. Also, because the scale of the image is increasing, more detail is present in the nonreference image that leads to a perturbation of the invariants. Additional smoothing of the reference and nonreference images reduces the perturbing effects of the scale change. Of the 12 matches 3 are incorrect and are eliminated by the gradient angle test.

The left image of Figure 2 shows frame 20 resampled into the image coordinate system of frame 01 in accordance with the global affine transformation robustly

estimated from nine matches. Resampling was by bi-linear interpolation. The right image of Figure 2 is the difference
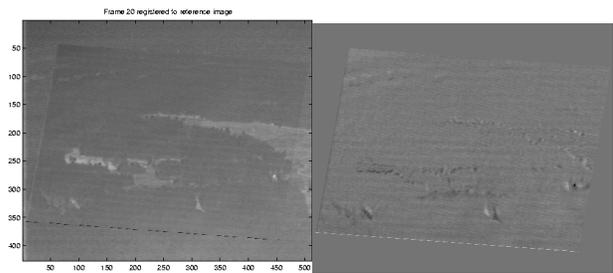
Figure 2: Left: Registered frame 20, compare with frame 01 in Fig. 1. Artifacts near image bottom are from source image resampled into registered frame. Right: Difference of reference and registered frames.

between the registered frame 20 and the reference frame 01. The contrast of the image is stretched to fill the range 0 to 255. The width of the brightness discontinuities are proportional to the error in the registration. The errors are due to the lack of match points in the foreground where perspective distortion is greatest. Also, the local distortion term has not been incorporated. Mislocalization of the corner points from Harris-Stephens corner detection can also contribute to the misregistration. The mislocalization is proportional to the filter scale of the derivative operators. An iterative registration refinement scheme may be a possible solution [3].
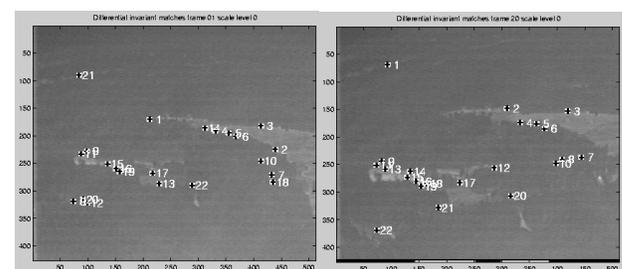
Figure 3: GDI matches for the first of three scales for infra-red DIM01 sequence. Left: frame 01. Right: frame 20.

4.2 Greylevel differential invariant matching extensions

Figure 3 shows the matches returned by the GDI matching method for the first of three scale levels with the differential invariant filter scale equal to 5 pixels. The remaining two reference scale levels are 5.5 and 6.05 pixels respectively. We note the larger number of matches, 22 in all. Of these 22 matches, 11 are correct or 50%.

Match 7 and 11 are correct but do not appear at all three scales. For the 5.5-pixel scale level the number of matches is 31 of which 14 are correct (45%), and for the last scale level there are 19 out of 29 matches correct (66%). With scale-space verification there are 9 out of 12

correct matches (75%) a large improvement in terms of the ratio of correct to total matches.
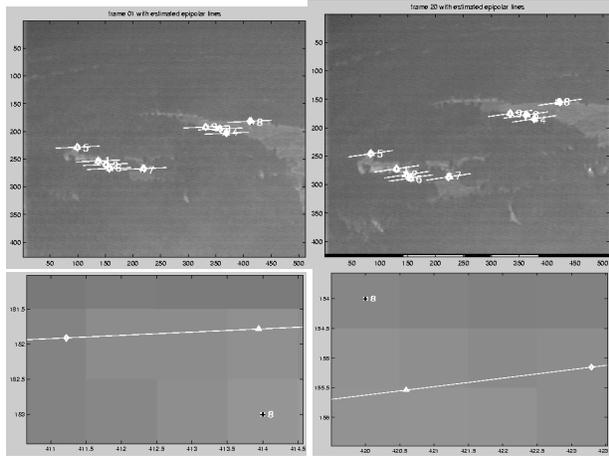
Figure 4. Epipolar lines for frames 01 and 20. Observations are marked with "+", reprojected scene points by "◊" and the height corrected reprojections by "Δ". The last two frames are a zoomed view of point 8 from frames 01 and 20 respectively.

We also note the high sensitivity to small perturbations of scale exhibited by the GDI matching process. Finally, the average number of records searched over both k-d trees over the three reference scale levels with 900 records in each tree was 76 in keeping with the expected logarithmic time to complete a query. Further results for these GDI matching extensions are reported in [10].
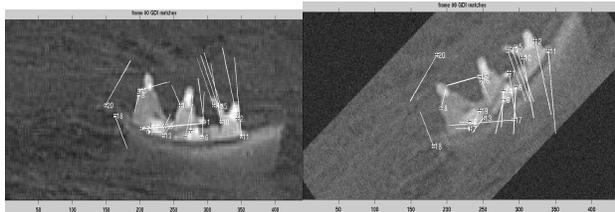
Figure 5: Reference image 00 with matches to frame 08 from GDI method. Matches 15, 17, and 20 are incorrect.

Figure 6: Labeled matches for frame 08 which is rotated 40º from the reference image with additive noise and a random affine distortion.

4.3 Epipolar estimation

The epipolar lines and height corrected reprojections were estimated according to Eqs. (8) and (9) and are plotted in Figure 4. The slope angles for the epipolar lines in frames 01 and 20 were –2.7° and –8.1° respectively for a fronto-parallel rotation angle of 5.4°.

The RMS residual for frame 01 reprojected points before height correction is 0.85 pixels and after height correction it is 0.52 pixels. The RMS residual for frame 20 is 0.97 and 0.69 pixels before and after height correction. The last two frames of Figure 4 shows point 8 with the fitted

epipolar line and the reprojected matching point from the other image before and after correction by the height local distortion factor.

4.4 Synthetic image sequence Lockheed Boats8

Next is an example of the registration process from a synthesized motion sequence. These images are from a sequence composed of 11 frames synthesized from a real infra-red image. The motion consists of a continuous rotation of the sequence reference image with additive Gaussian noise and a random affine distortion. In this example the non-reference image is frame 08, which is
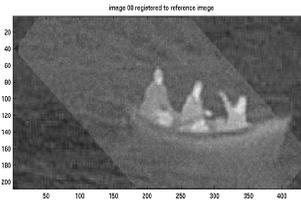
Figure 7: Frame 08 registered to reference image.
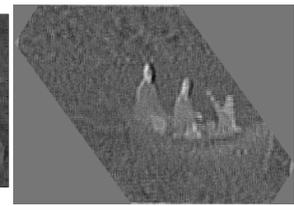
Figure 8: Difference between registered image 08 and the reference image. See text.

rotated by a 40º angle from the reference image. Figures 5 and 6 show the matches obtained from the greylevel differential invariant matching method. The GDI method provides 20 matches, where 17 are correct. Matches 15, 17, and 20 from Figures 5 and 6 are incorrect.
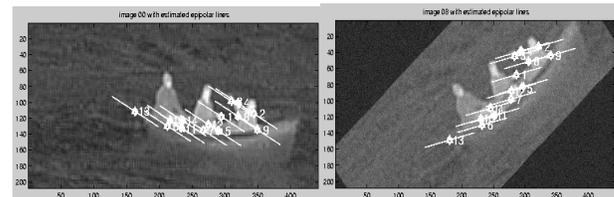
Figure 9. Lockheed sequence epipolar lines for frames 00 and 08.

Five matches, those labeled 3,8,15,16, and 20 from figures 5 and 6 were eliminated using the gradient angle test. Note that matches 3,8 and 16 were incorrectly eliminated. This may be due to the added noise and the shear introduced by the affine transformation in addition to the pure rotation. Figure 7 shows image 08 registered to the reference image after M-estimation of the global affine transformation and resampling by bi-linear interpolation. Incorrect match 17 from Figures 5 and 6 was correctly eliminated by the Tukey bi-weight estimation function, i.e., its weight was set to zero.

Figure 8 gives the image difference between the registered frame 08 and the reference image. The scene object can still be distinguished because the brightness is

not normally distributed after image subtraction in that area where the brightness values were clipped. Clipping occurred after adding zero mean noise with a standard deviation of 15 percent of the standard deviation of the reference brightness image. Figure 9 shows the estimated epipolar geometry for the two views. The epipolar line slope angle for frame 00 is 28.3º and –13º for frame 08 which gives a fronto-parallel rotation angle of 41.3º.

5   Conclusion

We have demonstrated the stabilization of aerial image sequences with scale, rotation and large translations using robust estimation. The method is computationally inexpensive and theoretically well founded owing to the appropriateness of the orthographic projection model which is a close approximation to the perspective projection for the aerial images. We have shown the greylevel differential invariant (GDI) method that provides matches for image transformations with rotation, scaling and some shear. Three extensions to the basic GDI matching method were described and experimentally verified. The extensions allow for runtime computation of the GDI vector normalization, increases the ratio of true to false matches and reduces the query time to an expected logarithmic time. We have applied the robust M-estimation to the solution of the registration function under the parallel projection model. We have shown the robustness of the method to outliers in the initial GDI matching.

In comparison, Zhang's method can provide many more matches than the GDI method in many cases, but cannot provide any correct matches when the affine transformation includes a large rotation or scaling. Therefore, the proposed approach is useful for aerial image registration and Zhang's method can be used for ground-based image registration. The initial matching sub-task, before epipolar verification, is clearly seen to be the major hurdle to a fully automatic feature-based registration method. New methods for this sub-task are presently under investigation.

Acknowledgements

References

[1]   Ballard, D.H., Wixson, L.E., "Object recognition using steerable filters at multiple scales," IEEE Workshop on Qualitative Vision, New York, NY, 2-10, June 1993.

[2]   Belanger, Louis N.; Gagnon, Roger; Lessard, Pierre; Maheux, Jean; Blanchard, Auguste, "Pixel and subpixel image stabilization at video rate for airborne long-range imaging sensors," in Proc. SPIE Vol. 2269, pp. 152-159, Infrared Technology XX, Bjorn F. Andresen; Ed., October 1994.

[3]   Bergen, J.R., Anandan, P., Hanna, K.J., Hingorani, R., "Hierarchical model-based motion estimation," European Conference on Computer Vision (ECCV 92), pp. 237-252, 1992.

[4]   Deriche, R., Zhang, Z., Luong, Q.-T., Faugeras, O., "Robust recovery of the epipolar geometry for an uncalibrated stereo rig," Proc. European Conference on Computer Vision, pp. 567-576, 1994.

[5]   Fischler, M.A., Bolles, R.C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Comm. of the ACM, Vol. 24, No. 6 , pp. 381-395, June 1981.

[6]   Harris, C., Stephens, M., "A combined corner and edge detector," Proceedings 4th Alvey Vision Conference, pp. 147-151, 1988.

[7]   Hu, X., Ahuja, N., "Feature extraction and matching as signal detection," International Journal of Pattern Recognition and Artificial Intelligence, 8(6), pp. 1343-1379, 1994.

[8]   Huber, P.J., "Robust Statistics," John Wiley & Sons, New York, 1981.

[9]   McReynolds, D.P., Lowe, D.G., "Rigidity checking of 3D point correspondences under perspective projection," IEEE Transactions PAMI,(18)12, pp. 1174-1185, 1996.

[10]   McReynolds, D.P., "Rigidity checking for matching 3D point correspondences under perspective projection," Ph.D. thesis, Dept. of Computer Science, University of British Columbia, 1997. ftp://ftp.cs.ubc.ca/pub/local/danm/Thesis

[11]   Pritt, M.D., "Image registration with use of the epipolar constraint for parallel projections," Journal of the Optical Society of America A, Vol. 10, No. 10, pp. 2187-2192, October 1993.

[12]   Schmid, C., Mohr, R., "Matching by local invariants," Rapport de recherche, N. 2644, INRIA, August 1995.

[13]   Schmid, C., Mohr, R., "Local Grayvalue Invariants for Image Retrieval," IEEE PAMI(19), No. 5, pp. 530-535, May 1997.

[14]   Schmid, C., Mohr, R., and Bauckhage, C., "Comparing and Evaluating Interest Points," Proc. International Conference on Computer Vision, pp. 230-235, Bombay, India, 1998.

[15]   Shapiro, L.S., Zisserman, A., Brady, M., "3D motion recovery via affine epipolar geometry," International Journal of Computer Vision, 16, pp. 147-182, 1995.

[16]   Sproull, R.F., "Refinements to nearest-neighbour searching in k-dimensional trees," Algorithmica, 6, 579-589, 1991.

[17]   Zhang, Z., "Parameter estimation techniques: a tutorial with application to conic fitting," Image and Vision Computing, 15, pp. 59-76, 1997.

[18]   Zhang, Z., Deriche, R., Faugeras, O., Quang-Tuan, L., "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," Artificial Intelligence, 78(1-2), pp. 87-119, 1995.