ROBUST IMAGE REGISTRATION UNDER SPATIALLY NON-UNIFORM BRIGHTNESS CHANGES

Dongjiang Xu and Takis Kasparis

Department of Electrical and Computer Engineering, Univ. of Central Florida
Orlando, Florida 32816, USA

ABSTRACT

Based on a non-linear image representation, we propose a novel robust approach for registration of images under spatial brightness changes. The image registration is formulated as a two-stage hybrid framework combining both a new point-based algorithm and a robust estimation with M-estimators in a coarse-to-fine manner. With the point-based algorithm applied at the highest level of decomposition, the initial affine parametric model can be first estimated. Subsequently, the robust estimation using M-estimators is incorporated into the hierarchical framework for completeness. Experimental results demonstrate that our proposed algorithm achieves higher accuracy and efficiency than the approach by brightness variation modeling with low-order polynomial functions (BVM).

1. INTRODUCTION

Image registration is to find the transformation which brings partially overlapped images into geometric alignment so that the points in one image can be related to their corresponding points in the other [1]. There are generally two types of distortions between the images to be registered. The first type is called a *spatial distortion*, where the images are spatially misaligned in relation to each other. The second type of distortions can be attributed to lighting conditions, weather, seasonal variations, etc., which is called a *non-spatial distortion*. The non-spatial distortions actually make the registration more difficult. Image registration techniques can be broadly classified into two categories: the intensity-based and feature-based methods. Intensity-based methods generally use the correlation function as a similarity metric, but correlation has less robustness for non-spatial distortions. In contrast, the features represent information on a higher level, which makes feature-based methods suitable for situations under non-spatial distortions. However, obtaining correct matches of features is sometimes a hard problem.

In this paper, we are concerned with the problem of registering the sensed image from the video camera attached to an aircraft with the reference imagery. We assume the distance between the camera and the target object on the ground is very far. For the domain of images under consideration, the variations of the intensity characteristics between the reference and sensed images may be large and non-uniform under non-spatial distortions. Motion estimation based on the optical flow equation (OFE) assumes brightness constancy [2][3]. If brightness is not conserved, the optical flow field can be severely biased approximation to the underlying 2D motion field of interest [4]. Researchers tried to relax this brightness constancy assumption and developed algorithms to estimate the motion parameters by the BVM-based method in an energy mini-mization framework [5][6]. It basically accounts for spatially varying smooth illumination variations. However, the situations resulting in the brightness changes between the reference and sensed images are very complex, and the effectiveness of BVM-based approach for image registration is sometimes limited. This prompts the necessity to identify an appropriate image representation, on which the OFE-based robust estimation using M-estimator in a coarse-to-fine manner can be incorporated. Our approach consists of two stages that address the difficulties caused by both spatial and non-spatial distortions. With a novel point-based algorithm applied at the coarsest scale of images, the initial affine model parameters can be estimated in the first stage. Subsequently, the robust estimation using M-estimator with a hierarchical iterative processing is used to refine the registration accuracy.

This paper is organized as follows. Section 2 describes the chosen image representation for extending OFE-based techniques. Section 3 describes a new point-based algorithm for initial matching. Section 4 describes the robust estimation using M-estimator. Section 5 describes the hierarchical algorithm and a selective data sampling strategy. Section 6 presents experimental results using real image pairs. Finally, we give our conclusions in Section 7.

2. THE IMAGE REPRESENTATION

OFE-based formulations assume brightness constancy, i.e., they estimate the 2D velocity of points of constant image brightness. However, the variations of the intensity characteristics between the reference and sensed images may be large and non-uniform because of non-spatial distortions. Therefore, in order to effectively incorporate the OFE-based parametric motion estimation into the proposed framework, two fundamental questions should be addressed: (1) what is a good image representation to work with using the OFE-based framework under non-spatial distortions; (2) the spatial distortions, i.e., the misalignment between images, may exceed certain large values, above which OFE-based methods can't converge to the correct result.

To capture the common intensity information while suppressing the non-common brightness changes, the transformation we have chosen is the absolute value of pixels in a

Laplacian pyramid, a non-linear image representation. The advantage of using the Laplacian pyramid is that its successive levels are band-passed versions of the original signal, which ensures that the low spatial frequencies containing the information about brightness changes are substantially removed. Non-linear band-pass representations are useful to image registration with both spatial and non-spatial distortions, because: (1) the creation of such representation images doesn't involve any thresholding, and therefore preserves all image details; (2) the image information removed in the creation of the non-linear representations is largely that which is not common to the two images due to brightness changes. This facilitates a coarse-to-fine search based on the non-linear image representation, where the method of modeling spatial brightness changes with low-order polynomial functions can't be directly applied.

3. A POINT-BASED ALGORITHM FOR INITIALIZATION

To avoid the problem of large spatial distortions, we propose a new point-based matching algorithm applied to the coarsest scale of both images for initial matching. With two point patterns extracted from the reference image, $A=\{a_1,...,a_m\}$, and the sensed image, $B=\{b_1,...,b_n\}$, respectively, the goal is to find the matching pairs between these two point sets. Only two matching pairs are required to determine the initial parameters.

For efficient matching, two relative invariant properties of similarity transform, the area and perimeter ratios, are used to characterize the mapping between the two point sets. Let R_{jk} be the area ratio, and P_{jk} the perimeter ratio, of a triangle pair from the reference and sensed images respectively.

$$R_{jk}=\gamma\log\left(S(\Delta_{1j})/S(\Delta_{2k})\right), j=1,...,C_3^m, k=1,...,C_3^n \tag{1}$$

$$P_{jk}=\beta\log\left(L(\Delta_{1j})/L(\Delta_{2k})\right), j=1,...,C_3^m, k=1,...,C_3^n \tag{2}$$

where $S(\Delta)$ and $L(\Delta)$ denote the area, and perimeter of a triangle respectively and γ and β are given constants. The proposed point-based matching algorithm consists of three steps: (i) geometric invariance properties between randomly selected triangles in the two images are evaluated; (ii) an accumulator is formed where votes on a particular match are tallied; (iii) corresponding point pairs are identified by a procedure of scanning the accumulator. In step (i), we create a 2D Area-Perimeter (*AP*) histogram based on the area and perimeter-length ratios. Based on the 2D histogram, all possible matching pairs of triangles can be identified. In step (ii), we form a matching Table $\{T_m(r,c)|r=1,...,m;c=1,...,n\}$ of control points. Any triangle pair $\{\Delta_{1j},\Delta_{2k}\}$ from the reference and sensed images respectively, corresponding to the maximum value of the *AP* histogram is selected as a candidate of a possible correct pair, where $\Delta_{1j}=(a_{1j_1},a_{1j_2},a_{1j_3})$ and $\Delta_{2k}=(b_{1j_1},b_{1j_2},b_{1j_3})$. Let $v_1=\left\|\overrightarrow{a_{1j_1}a_{1j_2}}\right\|/\left\|\overrightarrow{b_{1k_1}b_{1k_2}}\right\|$, $v_2=\left\|\overrightarrow{a_{1j_2}a_{1j_3}}\right\|/\left\|\overrightarrow{b_{1k_2}b_{1k_3}}\right\|$, $v_3=\left\|\overrightarrow{a_{1j_1}a_{1j_3}}\right\|/\left\|\overrightarrow{b_{1k_1}b_{1k_3}}\right\|$. If all the following conditions are satisfied:

$$|v_1-v_2|<\varepsilon_3,\quad |v_2-v_3|<\varepsilon_3,\quad |v_3-v_1|<\varepsilon_3 \tag{3}$$

where ε_3 is a small threshold, the three cells of T_m accumulate one vote, i.e., $T_m(r,c)=T_m(r,c)+1$, $r=j_1,j_2,j_3$, $c=k_1,k_2,k_3$, respectively. After Table T_m is formed, a "scanning algorithm" can determine the matching pairs. For the pairing procedure, first find the maximum value in each row. If that value is also the maximum in the corresponding column then we keep it and set all other values in the same row and column to zero. If not, then we set the entire row and column to zero. After this scanning, non-zero remaining values exceeding another threshold are used to determine the matching pairs. Finally, similarity transformation parameters estimated by the matching pairs can be used as the initial affine parameters in the second stage.

4. ROBUST ESTIMATION USING A DIRECT METHOD

In the proposed framework, the OFE is written as follows:

$$\Phi_x(x,y,t)u(x,y,t)+\Phi_y(x,y,t)v(x,y,t)+\Phi_t(x,y,t)=0 \tag{4}$$

where (u,v) is the motion vector and $\Phi(x,y;t)$ is the image function after non-linear transformation. The motion field can be described using a parametric model of a few parameters. The classical approach to motion estimation using OFE incorporates the motion model into OFE and establishes the objective function as:

$$E_D(u,v)=\sum_{\Re}\left((\nabla\Phi)^T\mathbf{u}+\Phi_t\right)^2 \tag{5}$$

where ∇I denotes the local brightness gradient vector, and $\mathbf{u}=[u,v]^T$ denotes the flow vector. In this paper, we focus on the estimation of an affine model for image registration. However, the proposed framework can be easily extended to other global models. Our robust formulation follows on the lines of standard M-estimation techniques [7][8], where the unknown parameters are estimated by minimizing an objective function of the residual error. In particular, the following minimization problem is solved:

$$\min_{\mathbf{a}} E_D(r;\sigma);\; E_D(r;\sigma)=\sum_{\Re}\rho\left((\nabla\Phi)^T\mathbf{u}(\mathbf{a})+\Phi_t;\sigma\right) \tag{6}$$

where $\rho(r;\sigma)$ is the robust ρ-function defined over the residuals, r; with a given scale factor, σ. In this work, we used the *Lorentzian* function, which is given as follows:

$$\rho(r;\sigma)=\log\left(1+\frac{r^2}{2\sigma^2}\right) \tag{7}$$

For the M-estimation for the parameters $\mathbf{a}$ in Eq. (6), we apply the Gauss-Newton (GN) method by Sawhney et al. [8]. Given a solution, $a^{(m)}$ at the m-th step, the descent direction, $\Delta a^{(m)}$, is given by

$$\Delta a^{(m)}=-H^{-1}(\mathbf{a}^{(m)})g(\mathbf{a}^{(m)}) \tag{8}$$

$$\mathbf{a}^{(m+1)}=\mathbf{a}^{(m)}+\lambda\Delta\mathbf{a}^{(m)} \tag{9}$$

for some positive λ. $H(\mathbf{a}^{(m)})$ is the approximation to the Hessian of the objective function in Eq. (6), involving only the first derivatives of the residuals, and $g(\mathbf{a}^{(m)})$ is its gradient, both defined at the current $\mathbf{a}^{(m)}$. We can write:

$$g_k=\sum_i\frac{\partial\rho}{\partial r_i}\frac{\partial r_i}{\partial\mathbf{a}_k}\qquad H_{kl}=\sum_i\frac{\partial^2\rho}{\partial r_i^2}\frac{\partial r_i}{\partial\mathbf{a}_k}\frac{\partial r_i}{\partial\mathbf{a}_l} \tag{10}$$

as the k-th and kl-th elements of g and H, respectively. Thus, the GN equations become,

$$\sum_l \sum_i \frac{\dot{\rho}(r)}{r_i} \frac{\partial r_i}{\partial a_k} \frac{\partial r_i}{\partial a_l} \Delta a_l = -\frac{\dot{\rho}(r_i)}{r_i} r_i \frac{\partial r_i}{\partial a_k} \qquad (11)$$

where $k, l = 1...K$, $i = 1...N$. It is noticed that the corresponding equations for the robust estimators are simply weighted normal equations with the weight of each measurement i being $\frac{\dot{\rho}(r_i)}{r_i}$. The minimization of the objective function assumes that the scale parameter σ is constant for each iteration. In this setting, the most commonly used scale estimator is the median absolute deviation (MAD) [9] estimate given by

$$\hat{\sigma} = 1.4825 \, \underset{i}{median}\left(\left| r_i - \underset{j}{median}(r_j) \right|\right) \qquad (12)$$

5. THE HIERARCHICAL ALGORITHM AND SELECTIVE DATA SAMPLING

Given the GN formulation and the step for σ estimation, we embed these in a hierarchical coarse-to-fine direct method. Starting at the coarse level, given an initial estimate of the parameters $\mathbf{a}^{(0)}$ from the initial matching, the iterations are repeated until the change in parameters is below a threshold or a specified number of iterations are reached. The estimated parameters are projected to the next finer level and used as initial estimates to warp the corresponding image, and the process repeated until convergence at the finest level.

In order to reduce the computational cost, a new selective data-sampling scheme is proposed, where only a sparse set of the pixels within the region of interest is used to form the objective function. In our selection scheme, we first partition the image into $m \times n$ uniform blocks. Then, we select a location in each block to form the objective function. For each pixel (x, y) in the block, we compute a local normalized-correlation surface $NC(x, y)$ around the displacement $\mathbf{u}(x, y; \mathbf{a})$ at the previous iteration. In our current implementation, the correlation surface is estimated only within a radius $d = 1$ around $\mathbf{u}(x, y; \mathbf{a})$. We define a reliability measure $\delta(x, y)$ at pixel (x, y) as the inverse of the sum of minimum distance errors in a quadratic model fitting to the local normalized-correlation surface $NC(x, y)$, i.e.,

$$\delta(x, y) = \frac{NC_x^2(x, y) + NC_y^2(x, y)}{e(x, y) + \varepsilon_2} \qquad (13)$$

where $e(x, y)$ is the sum of squared errors from the quadratic fitting in the local neighborhood of pixel (x, y), $NC_x(x, y)$ and $NC_y(x, y)$ the partial derivatives, and ε_2 a small positive number to prevent instabilities when $e(x, y)$ is very small. Thus, our selection of a reliable data constraint in each block is simply to find pixel (x, y) with the maximum reliability measure $\delta_{\max}$. From our experiment, we found 400 constraints are quite sufficient for accurate affine parameter estimation.

6. EXPERIMENTAL RESULTS

We have designed four sets of experiments: the first uses a pair of aerial images under irregular brightness changes, the second set of 4 image pairs (Test A-D) were captured with non-spatial distortions, the third set of 4 frame pairs (Test E-H) uses an aerial video sequence with small illumination changes, and the fourth consists of two pairs of Landsat TM images (Test I-J) with salient brightness changes. The size of the test image pairs is 512×512.

In the experiment, we limit our decomposition level to be 4, so that the coarsest image level uses 64×64, which still retains adequate information for feature extraction. The point-based matching algorithm can always identify enough matching pairs, and provide a good estimate, where the translation errors are within $\pm$ 10 pixels in both horizontal and vertical directions and $\pm$ 5 degrees for rotation errors. We first compare the registration accuracy between the proposed algorithm and the BVM-based method using an image pair of Fig. 1(a)-(b), and extracted points marked by black "+" in Fig. 1(c)-(d). A close examination of Fig. 2 indicates that the proposed algorithm greatly reduces the registration errors from the initial matching, whereas the motion estimate of the BVM-based is erroneous. For evaluation, we consider using the normalized cross correlation (NCC) between the overlapping areas of the images because the true motion is not known. From Table 1, it is noticed that the final accuracy of our algorithm is much better than that of the BVM-based approach for the second set of image pairs (Test A-D). For the third set of video frame pairs (Test E-F), the performances of both algorithms are nearly the same with only slight differences. For the fourth set of image pairs (Test I-J), the results are shown in Table 2, whereas the BVM-based method doesn't converge to the correct results in both pairs. For visual comparison, we selected two image pairs of Test A and I depicted in Fig. 3 and 4. It is observed the BVM-based method is not very robust when brightness variations between the reference and sensed images can't be modeled with low-order polynomial functions.

Table 1. NCC comparison.

	BVM	Our Approach
Test A	0.7803	0.9664
Test B	0.8308	0.9553
Test C	0.7643	0.9556
Test D	0.8575	0.9229
Test E	0.9722	0.9818
Test F	0.9745	0.9794
Test G	0.9783	0.9782
Test H	0.9717	0.9745

Table 2. NCC comparison using Landsat image pairs.

	Initial Matching	Our Approach
Test I	0.6280	0.9133
Test J	0.4871	0.8742

7. CONCLUSIONS

In this paper, we have proposed a robust hybrid hierarchical approach to the registration problem under spatially non-uniform brightness variations. Based on a non-linear band-pass image representation, the image registration is formulated as a two-stage procedure combining both the point-based algorithm and the robust estimation framework in a coarse-to-fine manner. As it is experimentally demonstrated, our proposed framework can achieve higher accuracy than the BVM-based approach.

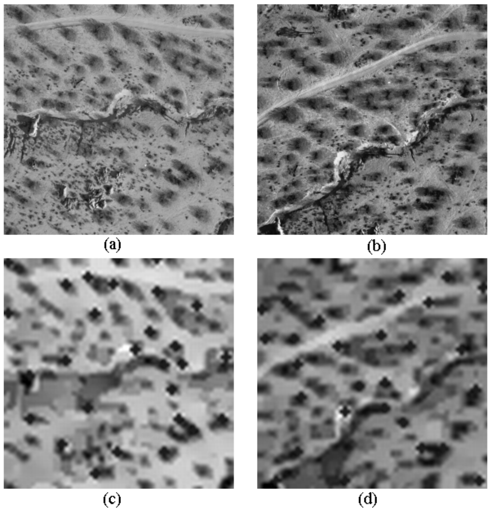

Fig. 1. Image registration example I: (a)-(b) two source images; (c)-(d) extracted points marked by black "+" at the coarsest scale.

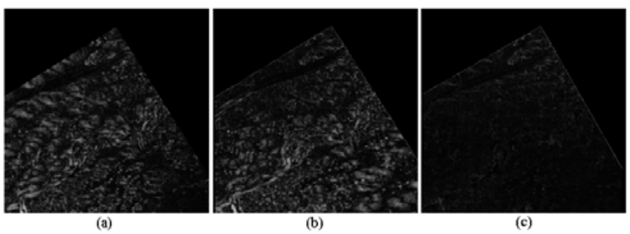

Fig. 2. Image difference: (a) after initial matching; (b) BVM-based method; (c) proposed approach.

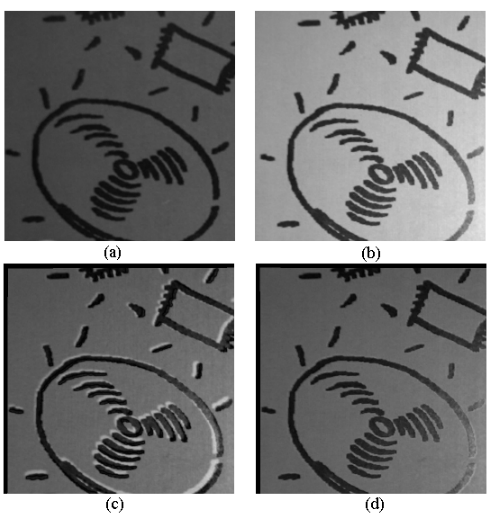

Fig. 3. Image registration example II: (a)-(b) two source images (Test A); (c)-(d) image differences using the BVM-based method, and the proposed approach, respectively.

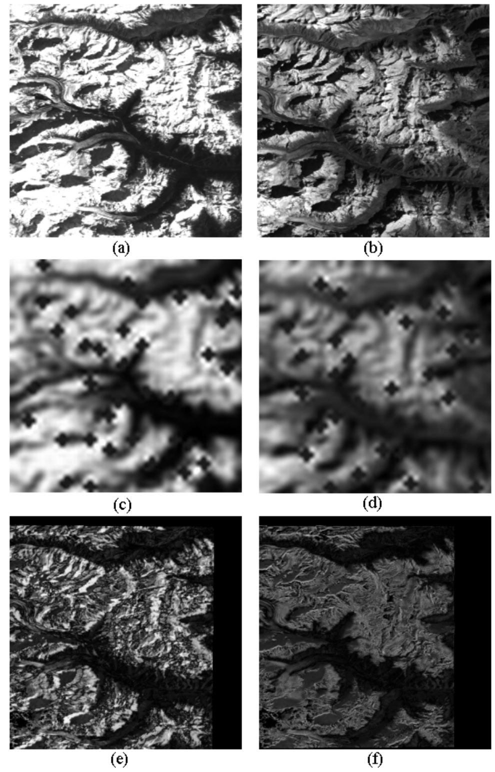

Fig. 4. Image registration example III: (a)-(b) two source images (Test I); (c)-(d) extracted points marked by black "+" at the coarsest scale; (e)-(f) image differences after the initial matching, and using the proposed approach, respectively.

8. REFERENCES

[1] L.G. Brown, "A Survey of Image Registration Techniques," *ACM Computing Surveys,* vol. 24, pp. 325-376, 1992.

[2] P. Anandan, "A Computational Framework and An Algorithm for Visual Motion," *IJCV*, vol. 2, pp. 283-311, 1989.

[3] J.R. Bergen, P. Anandan, K. Hanna, and R. Hingorani, "Hierarchical Model-based Motion Estimation," *Proc. 2nd ECCV*, pp. 237-252, 1992.

[4] S. Negahdaripour, and C. Yu, "A Generalized Brightness Change Model for Computing Optical Flow," *Proc. 4th ICCV*, pp. 11-14, 1993.

[5] S.-H. Lai, "Robust Image Matching under Partial Occlusion and Spatially Varying Illumination Change," *CVIU*, vol.78, pp. 84-98, 2000.

[6] Y. Altunbasak, R.M. Merserau, and A.J. Patti, "A Fast Parametric Motion Estimation Algorithm with Illumination and Lens Distortion Correction," *IEEE Trans. Image Processing*, vol.12, pp. 395-408, 2003.

[7] M.J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-smooth Flow Fields," *CVIU*, vol. 63, pp. 75-104, 1996.

[8] H.S. Sawhney, S. Ayer, and M. Gorkani, "Model-based 2D& 3D Dominant Motion Estimation for Mosaicing and Video Representation," *Proc. ICCV*, pp. 583-590, 1995.

[9] G. Li, "Robust Regression," in *Exploring Data Tables, Trends, and Shapes*, Wiley, New York, pp. 281-343, 1985.