FAST ROBUST PERSPECTIVE TRANSFORM ESTIMATION FOR AUTOMATIC IMAGE REGISTRATION IN DISASTER RESPONSE APPLICATIONS

Jim Thomas, Ahsan Kareem, Kevin Bowyer *

University of Notre Dame, Computer Science & Engineering, Notre Dame, IN 46556, USA

ABSTRACT

While automatic image registration has been extensively studied in other areas of image processing, it is still a complex problem in the framework of remote sensing for disaster response. This problem is difficult because there can be substantial change in the image content between the two images, and the time of day and lighting typically are different between the two images. In this work, we propose a two-step approach to achieve fast and robust registration of before- after-disaster aerial image pairs. First, the images are coarsely registered using a phase-correlation based algorithm. In the second step, transformed images are finely registered by matching features across grids and estimating the perspective transform. Our proposed algorithm is evaluated for robustness, accuracy and speed. It is found to achieve 100% registration success on 23 image pairs which proved challenging to either of the component approaches.

Index Terms— image registration, remote sensing, image processing, feature matching, phase correlation

1. INTRODUCTION

We address the problem of finding the perspective transform required for fine image registration. The satellite or aerial images from before and after a disaster may have been taken by different cameras at different altitudes, angles and positions. In addition, differences in camera parameters, local reflective properties of the object on the ground, changes due to storm damages and changes that occur over time may be present. Automatic registration is an important step in automated damage assessment from before- and after- storm images [1]. While there is no universal solution for image registration, this paper focuses on the combination of phase correlation with feature-based matching as a viable solution in the context of remote-sensing for change-detection applications. A comprehensive survey of general image registration methods was published in 2003 by [2]. Various area-based and feature-based registration algorithms were described and compared. A first comparison of image registration algorithms purely for remote sensing imagery was presented in [3]. They evaluated spatial correlation, phase correlation, iterative edge matching and wavelet maxima matching techniques for registration. These works focused on finding simple similarity transforms for registering multiview images. In contrast, our work focuses on pre- and post-even image pairs with a high degree of change and with perspective effects. We present a fast matching scheme that works well with multiple resolutions, relatively low overlap between images and substantial noise.

*The support for this study was provided in part by the Global Center of Excellence at TPU funded by the MEXT, Japan.

2. COARSE REGISTRATION

For coarse registration, we propose that perspective effects can be approximated using similarity transformation. Thus, approximate values for scale s, rotation θ and translation (t_x, t_y) can be estimated by applying phase correlation at a lower scale. Phase correlation relies on the translation property of the Fourier transform, which is referred to as the Fourier shift theorem. Let f_l and f_2 be the two images that differ only by a displacement (t_x, t_y) i.e.,

$$f_2(x, y) = f_1(x - t_x, y - ty) \tag{1}$$

Their corresponding Fourier transforms F_1 and F_2 will be related by

$$F_2(\xi, \eta) = e^{-j2\pi(\xi t_x + \eta t_y)} * F_1(\xi, \eta) \tag{2}$$

The cross-power spectrum of two images f_1 and f_2 is defined as

$$\frac{F_1(\xi, \eta)F_2^*(\xi, \eta)}{|F_1(\xi, \eta)F_2(\xi, \eta)|} = e^{j2\pi(\xi t_x + \eta t_y)} \tag{3}$$

where F_2^* is the complex conjugate of F_2, the shift theorem guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. By taking inverse Fourier transform of the representation in the frequency domain, we will have a function that is an impulse; that is, it is approximately zero everywhere except at the displacement (t_x, t_y) that is needed to optimally register the two images. The method can be extended to determine rotation and scaling differences between two images by performing phase correlation on the log polar images of the magnitudes of Fourier transforms [4]. However in practice, the determination of rotation and scaling was found to be unreliable and not robust to the changes between before- and after-storm images. Instead, we devise an algorithm that finds the scale and rotation which maximizes the peak of cross power spectrum. The proposed algorithm is detailed below.

1. *Downsample*: The images are smoothed and downsampled to 256×256 . This step is useful in reducing time required per each phase correlation step while ensuring a good approximation in computing similarity transformation.

2. *Preprocessing*: In practice, it is more likely that f_2 will be a linear shift of f_1, rather than a circular shift. In addition noise due to changes in the scene maybe present in f_2. In such cases, inverse Fourier transform of the cross power spectrum will not be a simple delta function, which will reduce the performance of the method. To overcome this, a hanning window function is applied to the images so that the edge effects can be ignored.

3. *Phase Correlation*: For $s_{min} \leq s \leq s_{max}$ and $\theta_{min} \leq \theta \leq \theta_{max}$, update $s = s + \delta_s$, $\theta = \theta + \delta_\theta$.

   Scale and rotate f_2 by s, θ to get f_2'. Perform phase correlation on f_2' and f_1, obtain the peak value r. Instead of looking for an interpolated peak, r is stored as the center of mass of the peak of the inverse Fourier transform of the cross power spectrum.

4. *Resampling*: The s, θ, t_x, t_y corresponding to the maximum value of r is used to compute a transformation function. f_2 is then resampled with this function resulting in a coarsely registered image pair.

3. SURF-BASED FEATURE EXTRACTION

For fine image registration, we adopt the use of feature-based matching and use blob detectors which are popular in object recognition. Our motivation for choosing this approach is that it is more robust than correlation-based techniques and works well even with low overlap percentage. In this approach, features are matched and a transformation model is calculated based on the matching. For feature extraction we use scale- and rotation-invariant interest point detector and descriptor, coined SURF (Speeded Up Robust Feature) first introduced in [5]. In SURF-based registration, firstly the interest points are selected at distinctive locations (mainly blobs) in the grayscale image. Next, the neighborhood of every interest point is represented by a feature vector. Finally, the descriptor vectors are matched between different images. The choice of SURF features for this application is inspired by the fact that they are more robust and faster than other state-of-the-art detectors and descriptors while maintaining high accuracy in matching.

4. INTEREST-POINT MATCHING

Once the interest points are detected and described, we need to find a one-to-one correspondence between the points detected from the two images. This is complicated by the fact that there may be outliers; that is, points found in either image that do not have a true corresponding point found in the other image. In naive nearest neighbor (NN) algorithm, each interest point in the test image is compared to each interest point in the reference image by calculating the Euclidean distance between their descriptor vectors. A matching pair is detected if distance of the nearest neighbor of an interest point is closer than the threshold $(t = 0.7)$ times the distance of the second nearest neighbor. Clearly, this is a $O(n^2)$ algorithm. Since the images are coarsely registered, the images are divided into $k \times k$ grids. Features points from corresponding grids are then matched.

Next, we need to find the perspective transform model that should be applied to register the images. Given a 3×3 perspective transformation matrix H with elements h_{ij} we determine the value of the matrix by minimizing the back projection error: $min \sum_i ((x_i' - (h_{11} * x_i + h_{12} * y_i + h_{13}) \div (h_{31} * x_i + h_{32} * y_i + h_{33}))^2 + (y_i' - (h_{21} * x_i + h_{22} * y_i + h_{23}) \div (h_{31} * x_i + h_{32} * y_i + h_{33}))^2)$. This is conventionally done by using a RANSAC algorithm [6] to try many different random subsets of 4 matched pairs each. For each subset, the back-projection error is calculated and the subset that produces the least error after a certain number of iterations is considered the correct one. The final perspective parameters are further refined with the Levenberg-Marquardt Algorithm.

4.1. Faster robust matching

The nearest neighbor (NN) search algorithm described previously is a slow $O(n^2)$ algorithm. This is especially true if the size k of the grids is large and many feature points need to be matched. Further, the ability of RANSAC algorithm to find a good subset depends on the number of inliers present in the matched point-pairs and termination criteria. A larger number of iterations in RANSAC can increase robustness at the cost of speed. To deal with the speed issues in the brute force NN matching, we explored various approximate NN matching algorithms.

Best bin first [7] is an approximate NN search algorithm that is based on a variant of kd-tree. In general, the algorithm returns the nearest neighbor for a large fraction of queries and a very close neighbor otherwise. We also compared two other approximate NN algorithms introduced in [8] , namely randomized kd-trees and hierarchical k-means trees. These algorithms were evaluated for their execution times. This is reported in Figure 1. A speed-up of 30x was found at finer resolutions (and consequently larger image sizes) for approximate NN matching algorithms.

To deal with the problem of reduced inliers in approxi-

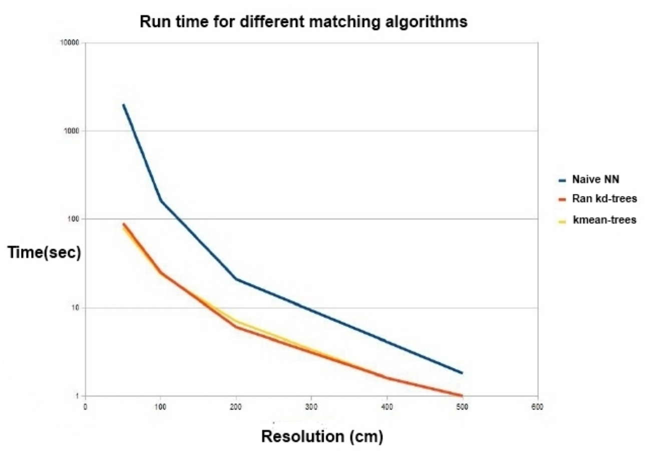

Fig. 1. Elapsed real time for matching points at different resolutions using randomized kd-trees, heirarchial k-means trees and Naive NN search. The algorithms were run on a laptop with Intel core i5 2.4 Ghz processor.

mate NN matching, we propose a modified form of RANSAC similar to CONSAC described in [9]. The proposed constrained form of RANSAC is different in the following ways: a) When subsets of 4 matched pairs were chosen, these 4 pairs had to follow certain geometric constraints; namely, cyclic or anti-cyclic order had to be conserved and collinearity (of any 3 points) had to be conserved. b) The upper limit on the number of iterations of RANSAC was determined by a greedy method which counts the number of iterations since last time the best subset was updated. An unregistered image-pair and the transformed pair after registration using the proposed algorithm are shown in Figure 2.

5. EVALUATION

To study the effect of image resolution on the proposed registration process, we used images from publicly available NOAA and USGS aerial imagery dataset. This included 15 image pairs from different locations in coastal Florida before- and after- Hurricane Dennis. Additionally, we used 8 image pairs from Texas before and after Hurricane Ike. These images were 4077×4092 and 50cm resolution originally. They were downsampled to smaller versions with 1m, 2m, and 4m resolutions. The image pairs vary in overlap, lighting conditions and viewpoint. Ground truth for registration was prepared by manually marking 25 corresponding landmark points for each image in a pair. Registration error is calculated by taking the average distance between the marked points and the transformed points.

We evaluated phase-correlation, feature-driven registration, and hybrid phase-feature registration. Proposed methods of evaluation include registration success, execution time and registration error. Our metric for successful registration is the percentage of images that are deemed to be registered with less than a 6-pixel error in the finest resolution. The effect of registering at various resolutions is shown in Table 1. The average registration errors and run times are shown in Table 2. The average error figures presented represent only the image pairs that registered with less than 50px error.

Among feature-driven methods, we evaluated 2 matching schemes. They included Naive NN with RANSAC and approximate NN with the proposed modified CONSAC. The phase-correlation algorithm presented in Section 2 was evaluated with $s_{min} = 0.7$, $s_{max} = 1.2$, $\theta_{min} = 0$, $\theta_{max} = 360$, $\delta_s = 0.1$ and $\delta_\theta = 1$. All the images were registered with about 18px error and 22% of the image-pairs with less than 6px error with phase-correlation. In Table 1, while approximate NN with modified CONSAC performed better than the naive approach, registration using phase correlation had a lower success rate indicating its suitability only for a coarse registration.

Our proposed hybrid phase-feature method combines phase-correlation for coarse registration with feature-based matching for fine registration. The new algorithm performed with 100% registration success on our dataset. Also, the average registration error and run time both turn out best for our proposed approach. Using this scheme, images are registered with 3 pixel error or less and 4x total speed up in comparison to feature-based matching without a phase-correlation step.

Table 1. REGISTRATION SUCCESS AT VARYING IMAGE RESOLUTIONS

Algorithm	400cm	200cm	100cm	50cm
Naive NN and RANSAC	22%	35%	74%	74%
Approximate NN and modified CONSAC	65%	78%	78%	78%
Phase Correlation	22%	22%	22%	22%
Phase Correlation, Approximate NN on grids and modified CONSAC	100%	100%	100%	100%

6. CONCLUSIONS

This work presented an automatic registration scheme which is fast and robust enough to work well with real world remote sensing imagery. From the study described in the previous section, it is clear that a combination of coarse registration using phase correlation and fine registration with an approximate nearest neighbor search algorithm combined with a constrained RANSAC algorithm for point-pairs subset selection registers the images with 4x speed-up and least regis-

Fig. 2. A sample image pair of before (top) and after (bottom) registration. This pair is of a region of Florida before- and after-Hurricane Dennis.

Table 2. AVERAGE REGISTRATION ERROR AND EXECUTION TIME AT 50CM RESOLUTION

Algorithm	Error	Time
Approximate NN and modified CONSAC	9px	465sec
Phase Correlation	18px	42sec
Phase Correlation, Approximate NN on grids and modified CONSAC	3px	126sec

tration error. This is done without compromising robustness and is hence recommended for change detection and disaster response applications.

7. REFERENCES

[1] J Thomas, K W Bowyer, and A Kareem, "Towards a robust automated hurricane damage assessment from high-resolution images," in *13th International Conference on Wind Engineering*, 2011.

[2] B Zitova and J Flusser, "Image registration methods: a survey," in *Image and Vision Computing*, 2003, pp. 977–1000.

[3] JL Moigne, W Xia, and P Chalermwat, "First evaluation of automatic registration methods," in *IEEE International Geoscience and Remote Sensing*, 1998.

[4] B S Reddy and B N Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," in *IEEE Transactions on Image Processing,*, 1996, vol. 5, pp. 1266–1272.

[5] H Bay, T Tuytelaars, and L V Gool, "Surf: Speeded up robust features," in *European Conference on Computer Vision*, 2006, pp. 404–417.

[6] Z Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *IEEE International Conference on Computer Vision*, 1999, vol. 1, pp. 666–678.

[7] J Beis and D G Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Computer Vision and Pattern Recognition*, 1997, pp. 1000–1006.

[8] M Muja and D G Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application*, 2009, pp. 331–340.

[9] F Guo, G Aggarwal, K Shafique, X Cao, Z Rasheed, and N Haering, "An efficient data driven algorithm for multisensor alignment," in *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, 2008.