Proceedings of the 2008 IEEE
International Conference on Information and Automation
June 20 -23, 2008, Zhangjiajie, China

Aerial Image Registration based on Joint Feature-Spatial Spaces, Curve and Template Matching

Guifeng Shao and Fenghui Yao
Department of Computer Science
Mohan J. Malkani
Department of Electric and Computer Engineering
Tennessee State University
3500 John A Merritt Blvd, Nashville, TN 37215, USA
{gshao, fyao, mmalkani }@tnstate.edu

Abstract - **Image registration has wide applications in remote sensing, medicine, cartography, and computer vision. This paper describes a method for aerial image registration, which is based on joint feature-spatial spaces, curve matching, and template matching. The entire algorithm consists of (i) segmentation and region representation, (ii) region matching based on joint feature-spatial space, (iii) transformation model estimation based on curve matching and template matching, and (iv) image transformation. Experiment results show the effectiveness of this algorithm.**

I. INTRODUCTION

Image registration is a process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It has wide applications in remote sensing (multispectral classification, environmental monitoring, geographic information system (GIS), change detection, image mosaicing, weather forecasting, and creating super-resolution images), in medicine (combining computer tomography (CT) and Nuclear Magnetic Resonance (NMR) data to obtain more complete information about patient, monitoring tunor growth, and treatment verification), in cartography (map updating), and in computer vision (target localization, automatic quality control). In last ten years, thousands of papers were published on the topic of image registration. The majority of the registration methods consist of the following four steps [1]: (1) *feature detection*, (2) *feature matching*, (3) *transform model estimation*, and (4) *image resampling and transformation*. In the first step, salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or, preferably, automatically detected. For the further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points). In feature matching step, the correspondence between the features detected in the sensed image and those detected in the reference image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose. In the third step, the type and parameters of the so-called mapping functions, aligning the sensed image with the reference image, are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence. In the last step, the sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique. Features for image registration can be point features, line features, or region features. Features detected in the reference and sensed images are matched by means of the image intensity values in their close neighborhoods, the feature spatial distribution, or the feature symbolic description. These methods can be classified into two categories, area-based and feature-based methods, respectively. Area-based methods contain correlation-like methods, Fourier methods, mutual information methods, and optimization methods. Feature-based methods contain spatial relation methods, invariant descriptor methods, relaxation methods, and methods using pyramids and wavelets. Feature-based methods are typically applied when the local structural information is more significant than the information carried by the image intensities. They allow to register images of completely different nature (like aerial photograph and map) and can handle between-image distortions. The drawback of the feature-based methods is that the respective features might be hard to detect and/or unstable in time. Area-based methods are preferably applied when the images do not have many prominent details and the distinctive information is provided by grey-levels/colors rather than by local shapes and structure. Area-based methods have two principal limitations. Reference and sensed image must have somehow similar intensity functions, either identical or at least statistically dependent. From the geometric point of view, only shift and small rotation between the images are allowed when using area-based method.

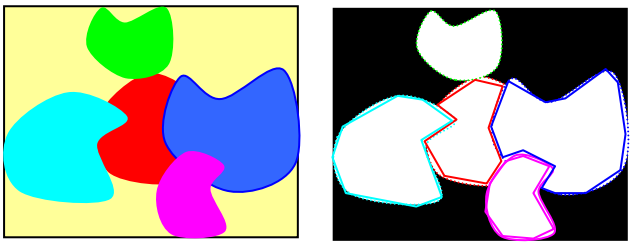
This work focuses on aerial image registration, in which the images are recorded by a camera mounted on an UAV (unmanned aerial vehicle). UAV hover motion contains roll-, pitch-, and yaw-rotation, and translation. The UAV image sequence contains large shift and large rotation, so that the respective features cannot be detected stably. Therefore, feature-based methods and area-based methods do not work well, because of their drawbacks as mentioned above. To remedy the drawbacks of feature-based methods and area-based methods, this paper proposes a method based on joint feature-spatial spaces, curve matching, and template matching. The proposed method belongs to the area-based method category. It is different with the previous methods in: (i) it employs the joint feature-spatial spaces of the region; (ii) the sample points can be randomly chosen; (iii) region contours and template matching are employed to estimate the

978-1-4244-2184-8/08/$25.00 © 2008 IEEE.

transformation model. Aerial images are employed to test the algorithm. Experiment results verify the effectiveness of the proposed algorithm.

II. Proposed Algorithm

Our algorithm for aerial image registration can be formulated in four steps: (i) segmentation and region polygon representation, (ii) region matching based on joint feature-spatial space, (iii) estimation of transformation model, and (iv) image transformation. These steps are explained in greater detail in the following.



(a) Segmentation map (b) Contours and approximate polygons

Fig. 1 Generation of image segments and their polygon approximation.

A. Segmentation and Region Polygon Representation

This preprocessing step is illustrated in Fig. 1. Let F_m represent the m-th input image. A spatial segmentation of F_m is obtained by using JSEG [2] algorithm. JSEG algorithm generates a set of non-overlapping regions. Let $R_m = \{r_{m1}, r_{m2}, \ldots, r_{mK}\}$ represent these regions, where K is the number of regions (refer to Fig.1 (a)). Contours for all regions are detected by applying contour detection algorithm. Let $B_m = \{b_{m1}, b_{m2}, \ldots, b_{mK}\}$ represent the contours detected from the regions (refer to dotted-curves in Fig.1 (b)). For regions, whose lengths are smaller than contour threshold T_{con} (currently, it is set at 50), are considered as noises, and are filtered out. For regions, whose circumscribed rectangle size are larger than the rectangle threshold T_{rect} (currently, it is set at $0.8\times$ H, where H is the height of F_m), are considered as background area, and are removed. To reduce the computation time in the succeeding steps, contours are approximated with polygons by applying Douglas-Peucker algorithm [3]. The obtained polygons are denoted by $P_m = \{p_{m1}, p_{m2}, \ldots, p_{mK}\}$ (see solid-polygons in Fig.1 (b)). Fig. 2 (a) and (b) show two input images, the segmentation results are shown by green contours, and their approximate polygons are marked by purple line. Each closed contour represents a image region. Each image region's circumscribed rectangle is calculated from its contour coordinates. The image region whose max{width, height} is greater than H/2 is filtered out. After this filtering process, six image regions in input image in Fig. 2 (a) are left, which are numbered from 0 to 5, and seven image regions left for the input image in Fig. 2 (b), which are numbered from 0 to 6. These image regions will be used for similarity measure calculation in the next section.

B. Joint Feature-Spatial Spaces and Similarity Measure

Let u_i be a d-dimensional feature vector at image location x_i, and $S=\{y_i = (x_i, u_i)\}$ (i = 1, …, N) be samples from a image region in $r_{mk} \in R_m$ (k = 1, 2, …, K). The estimate of the probability at (x, u) in joint space is:

$$\hat{P}(x,u) = \frac{1}{N}\sum_{i=1}^{N} K_\sigma(x-x_i)G_h(u-u_i) \tag{1}$$

where K_σ is a 2-dimensional kernel with a bandwidth σ, and G_h is a d-dimensional kernel with a bandwidth h. The bandwidth in the spatial dimensions represents the variability in feature location due to the local deformation or measurement uncertainty, while the bandwidth in the feature dimensions represents the variability in value of feature.

Given two distributions with samples $I_x = \{(x_i, u_i)\}$ (i = 1, …, N) and $I_y = \{(y_j, v_j)\}$ (j = 1, …, M), the similarity measure between I_x and I_y is defined as:

$$\begin{aligned} J(I_x, I_y) &= \frac{1}{M}\sum_{i=1}^{M}\hat{P}_x(y_j, v_j) \\ &= \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M} K_\sigma\left(\left\|\frac{x_i - y_j}{\sigma}\right\|^2\right) G_h\left(\left\|\frac{u_i - v_j}{h}\right\|^2\right). \end{aligned} \tag{2}$$

$J(I_x, I_y)$ is symmetric and bounded by zero and one. This similarity is based on the average separation criterion in cluster analysis [4] except that it employs the distance with a kernelized one. This similarity measure has been applied for target tracking [5] [6].

For the sensed image F_m and reference image F_n, the region matching process is as follows.

For two image regions, $r_{mk} \in R_m$ and $r_{nl} \in R_n$ (k = 1, 2, …, K, l = 1, 2, …, L, and R_m and R_n are region sets obtained from image F_m and image F_n, respectively, according to the processing described in 2.1), if they satisfy

$$\begin{cases} \left\|C(r_{mk}) - C(r_{nl})\right\| < H/4 \\ \dfrac{\left|L(r_{mk}) - L(r_{nl})\right|}{\min\{L(r_{mk}), L(r_{nl})\}} < L_{thres} \\ \left|S(r_{mk}) - S(r_{nl})\right| < S_{thres} \end{cases}, \tag{3}$$

the similarity measure, J_{mn}^{kl}, between them is calculated according to equation (2), otherwise J_{mn}^{kl} is set to zero, where $C(r_{ij})$, $L(r_{ij})$, and $S(r_{ij})$ is the center, perimeter, and shape of the region r_{ij}, respectively, and $S(r_{ij})$ is given by $\dfrac{\pi L^2(r_{ij})}{4A^2(r_{ij})}$

Fig. 2 Image segmentation, curve matching, and image registration results. (a) First input image (frame 19); (b) Second input image (frame 20); (c) Curve matching; (d) Image registration result.

and $A(r_{ij})$ is the area of the region r_{ij}. Then the maximal similarity measure, $J_{mn}^{\hat{k}\hat{l}}$, is obtained as $J_{mn}^{\hat{k}\hat{l}} = \max\{J_{mn}^{kl}\}$ for k = 1, …, $\hat{k}$ …, K, and l = 1, …, $\hat{l}$, …, L). This means that two images F_m and F_n overlaps at $\hat{k}$-th region in F_m and $\hat{l}$-th region in F_n. Note that $C(r_{ij})$, $L(r_{ij})$, $S(r_{ij})$ and $A(r_{ij})$ are obtained by employing contour coordinates and L_{thres} and S_{thres} are threshold values which are determined experimentally. As shown in Fig. 2 (c), the image region number 4 in input image (a) matches the image region number 5 in (b). The matched regions will be used to estimate the transformation model between two images.

C. Estimation of Transformation Model

The relationship between the sensed image point (X, Y) and the reference image point $(\hat{X}, \hat{Y})$ is modeled by 2-D Affine transform:

$$\begin{pmatrix} \hat{X} \\ \hat{Y} \end{pmatrix} = \kappa \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \Delta X \\ \Delta Y \end{pmatrix}, \quad (4)$$

where $(\kappa, \varphi, \Delta X, \Delta Y)$ corresponds to scaling factor, a rotation angle, and translation along the two orthogonal directions, respectively. These parameters are reasoned in the following way. For the time being, κ is set at one. For $\hat{k}$-th region in F_m (note that it is the $\hat{k}$-th region in F_m, that matches with $\hat{l}$-th region in F_n), its corresponding polygon is $p_{m\hat{k}} \in P_m$. Let $V_{m\hat{k}} = \{v_{m\hat{k}}^1, \ldots, v_{m\hat{k}}^U\}$ and $\theta_{m\hat{k}} = \{\theta_{m\hat{k}}^1, \ldots, \theta_{m\hat{k}}^U\}$ represent vertices

and vertex angles of polygon $p_{mk\hat{}}$, where U is the number of vertices of $p_{mk\hat{}}$, and $\theta^i_{mk\hat{}}$ is the angle formed by two line segments, one from (i-1)-th vertex to i-th vertex and another one from i-th to (i+1)-th vertex, when tracing vertices clockwise. Let $S_{mk\hat{}} = \{ s^1_{mk\hat{}}, \ldots, s^U_{mk\hat{}} \}$ represent the shape of polygon $p_{mk\hat{}}$ at each vertex. $s^i_{mk\hat{}}$ (i = 1, 2, …, U) takes the value zero for concave vertex, and one for convex vertex. The polygon $p_{mk\hat{}} \in P_m$ in F_m and $p_{nl\hat{}} \in P_n$ in F_n are matched at $v^i_{mk\hat{}} \in p_{mk\hat{}}$ and $v^j_{nl\hat{}} \in p_{nl\hat{}}$ if they satisfy:

$$\begin{cases} \theta^i_{mk\hat{}} = \min\{\theta^1_{mk\hat{}}, \ldots, \theta^U_{mk\hat{}}\} \text{ and } \theta^j_{nk\hat{}} = \min\{\theta^1_{nk\hat{}}, \ldots, \theta^V_{nk\hat{}}\} \\ \left|\theta^i_{mk\hat{}} - \theta^j_{nk\hat{}}\right| < \theta_{thres} \text{ and } \left|\theta^{i-1}_{mk\hat{}} - \theta^{j-1}_{nk\hat{}}\right| < \theta_{thres} \text{ and } \left|\theta^{i+1}_{mk\hat{}} - \theta^{j+1}_{nk\hat{}}\right| < \theta_{thres} \\ s^i_{mk\hat{}} = s^j_{nk\hat{}} \text{ and } s^{i-1}_{mk\hat{}} = s^{j-1}_{nk\hat{}} \text{ and } s^{i+1}_{mk\hat{}} = s^{j+1}_{nk\hat{}} \end{cases} \tag{5}$$

where V is the number of vertices of $p_{nl\hat{}} \in P_n$ and θ_{thres} is a predetermined threshold value. Then $(\Delta X, \Delta Y) = v^i_{mk\hat{}} - v^j_{nl\hat{}}$ (Note that vertex is 2-D vector).

After translation parameters are determined, the vertex $v^j_{nl\hat{}}$ in image F_n is overlapped at $v^i_{mk\hat{}}$ in image F_m by translating F_n. And then image F_n is rotate in the range from $-\delta_{thres}$ to δ_{thres} with the increment of 1° in each repetition. The match error at rotation angle δ is calculated as,

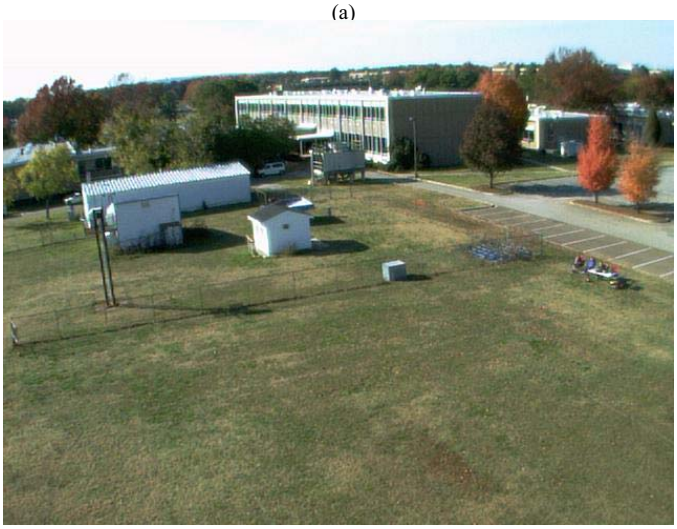
$$E_\delta = \frac{\sum_{x,y\in R} I_{F_m}(x,y) I^{rot}_{F_n}(x,y)}{\sqrt{\sum_{x,y\in R} I_{F_m}(x,y)^2 \sum_{x,y\in R} I^{rot}_{F_n}(x,y)^2}}, \tag{6}$$

where R is the template area centered at $v^i_{mk\hat{}}$, and $I^{rot}_{F_n}(x,y)$ is the template image from rotated image of F_n. Image F_m and F_n are best matched at angle φ where $E_\varphi = \min\{E_\delta\}$, $\delta \in [-\delta_{thres}, \delta_{thres}]$. These Affine parameters are passed to the next stage for image transformation. Fig. 2 (c) shows that polygons for the matched image region in image (a) and (b) are overlapped at the vertex marked by white rectangle, and the red polygon in image (b) is rotated to determine the rotation angle. The translation and rotation parameters are used in image transformation in next section.

D. Image Transformation

The processing in this step contains image translation, rotation, and interpolation. At beginning of image registration, the registration image F_R is initialized by zero. Note that F_R must be large enough to register all input images. F_m is copied to F_R. And then F_n is translated and rotated, and stitched to F_R. For the succeeding images in image sequence, the registration

(a)



(b)

(c)

Fig. 3 Image registration result. (a) First input image (frame 23); (b) Second input image (frame 24); (c) Curve matching and image registration result.

process is to repeat the above operation. Fig. 2 (d) shows the image registration result for the input image in (a) and (b).

III. Experiment Results

The proposed algorithm is coded by using MS Visual C++ 6.0 and OpenCV [7], running on Windows platform. Threshold values L_{thres} and S_{thres} in equation (3) are set at 0.5 and 0.2, respectively, θ_{thres} , R, δ_{thres} in equation (5) and (6) is set 30°, 120×120 pixels, and 5°, respectively. All input images are 640×480 color images. Fig.3 (a) and (b) show two input images (frame 23 and 24), and (c) shows the curve matching and image registration results. Fig. 4 (a) shows four input images from frame 19 to frame 22, Fig. 4 (b) shows the image registration result. The experiments are conducted on Windows Vista machine mounted with an Intel Core 2 CPU and 2GB memory, running at 2.33GHz. The average processing times for each processing step are shown in Table 1. Note that each registration test employed two images. Total processing time for registering two images is 89 seconds.

TABLE I PROCESSING TIME (ms)

Processing Task	Processing Time
Image segmentation	87,033 (97.2%)
Region matching	515 (0.6%)
Affine transform Estimation	1,804 (2.0%)
Image Registration	144 (0.2%)
Total	89,495

IV. Conclusions and Future Works

This paper described an aerial image registration algorithm which is based on joint feature-spatial spaces, curve matching, and template matching. The entire algorithm consists of (i) segmentation and region representation, (ii) region matching based on joint feature-spatial space, (iii) transformation model estimation based on curve matching and template matching, and (iv) image transformation. Experiments were conducted by employing images taken from UAV (unmanned aerial vehicle) flying at height of about 10 meter. Experiments results show the effectiveness of this algorithm.

Image segmentation step takes 97.2% of the entire processing time. This can be improved by using computer cluster. Currently, each segment is approximated by a polygon. Polygon region matching brings about large errors if image is not properly segmented, and it also brings mistakes. The polygons for region representation need to be triangulated to do precise matching.

This paper discussed the image registration for aerial images collected by an UAV (unmanned aerial vehicle). This approach can be applied to the terrain map generation and environment recognition for UGVs (unmanned ground vehicles) to conduct terrain mapping, obstacle detection and avoidance, and autonomous navigation, by mounting a pan-tilt-zoom camera on the top of UGS. All these are left to do in the future.

Acknowledgement

The aerial video data is provided by Dr. Jeffery Cerny, at Rotorcraft Systems Engineering and Simulation Center, the University of Alabama in Huntsville.

References

[1] B. Zitova and F. Flusser, "Image Registration methods: A Survey," *Image and Vision Computing,* 21, pp. 977-1000, 2003.

[2] Y, Deng and B.S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Video," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 23, pp. 800 -810, No.8, August 2001.

[3] D. Douglas and T. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *The Canadian Cartographer* 10(2), pp. 112-122, 1973.

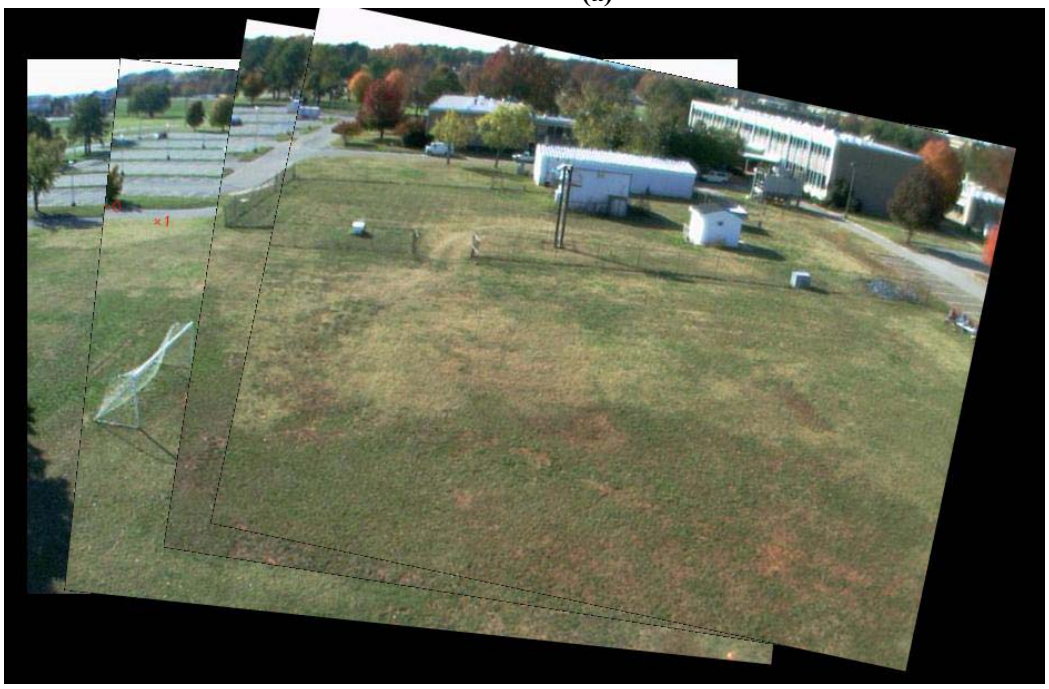
[4] A. R. Webb, "Statistical Pattern Recognition," John Weley & Sons, UK, 2nd Edition, 2002.

[5] A. Elgammal, R. Duraiswami, and L. S. Davis, "Probabilistic Tracking in Joint Feature-Spatial Spaces," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Wisconsin, USA, June 16-22, 2003.

[6] C. Yang, R. Duraiswami and L. Davis, "Efficient Mean-Shift Tracking via a New Similarity Measure," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, USA, June 20-25, 2005.

[7] http://sourceforge.net/project/showfiles.php?group_id=22870&package_id=16937

(a)



(b)

Fig. 4 Image registration results. (a) Four input images (frame 19 to 22); (b) Image registration result.