ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 5, May 2016

Multimodal image registration of UAV and Satellite images using Mutual Information

G Ellakiya[1], Deeksha K[2], Rekha R[3], Naziya Begum[4],
UG Student [ECE], Department of ECE ,M S Engineering College, Bangalore, Karnataka, India 1,2,3,4

Parinitha J[5]
Assistant Professor, Department of ECE ,M S Engineering College, Bangalore, Karnataka, India

Abstract: In this paper we have developed a multimodal image registration system for Unmanned Aerial Vehicle (UAV) image and Google Satellite image. This method uses a small number of automatically extracted scale invariant salient region features, whose interior intensities can be matched using robust similarity measures and mutual information to refine registration between UAV image and Google satellite image. The step involved in this method are orthorectification, feature detection, feature matching, RANSAC (model fitting) and transformation. This method gives more accuracy than the conventional image registration methods.

***Keywords*—** Orthrectification, Salient region features, RANSAC, Image registration, UAV image, Satellite image

I. INTRODUCTION

Image registration is the process of aligning two or more images of the same scene. This process involves designating one **image** as the reference (also called the reference **image** or the fixed **image**), and applying geometric transformations to the other images so that they align with the reference.[1]

Image registration is the process of aligning two or more images of the same scene and applying geometrical transformation that aligns points in one view of an object with corresponding Image registration responding points in another view of that object or another object.Data may be multiple photographs, data from different sensors, times, depths, or viewpoints has adaptability, computing speed and registration accuracy[7].

An UAV (Unmanned Aerial Vehicle),known as drone is an aircraft without a human pilot . (Capability to fly at altitudes of 30000 ft with payloads.) Its flight is either controlled autonomously by computers or under the remote control of a pilot on the ground. They are equipped with a variety of sensors, such as SAR imaging modes, E/O(electro optical sensor) and IR sensor technology.

Registration requires two inputs satellite image as reference image and UAV image to be registered. Multimodal image registration can be viewed as the task of integrating information from multiple sources. It consists of transformation of the overlaid image to the reference image coordinate system by means of appropriate transformations. Those transformations are based on specific characteristics of the multimodal data.

II. PREVIOUS WORK

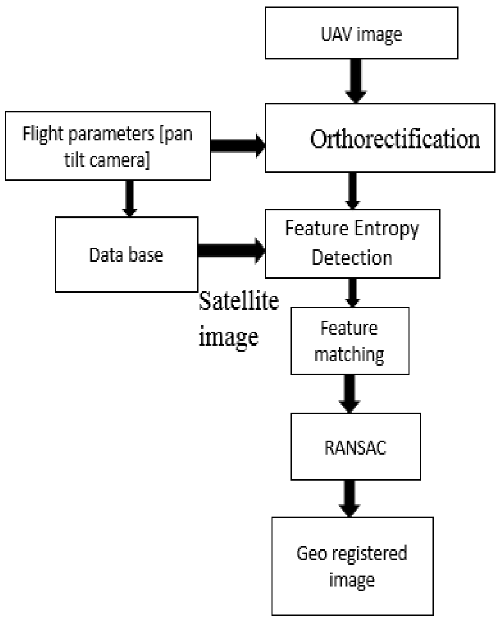
The proposed image registration method is largely inspired by the pioneering works from the object recognition literature[2,3]. From their works, we learned two important aspects that would be beneficial when used in image registration. The first aspect is the use of scale-invariant *region* features. In [2], objects are modeled as flexible constellations of regions (parts) in order to learn and recognize object class models. An entropy-based feature detector [3] is used to select region features that have complex intensity distributions and are stable in both spatial and scale spaces. When adapting the idea of region features to solve image registration problems, suitable and robust similarity measures need to be defined between region intensity values, to deal with multi-modal matching, image noise, and intensity inhomogeneity. The second aspect is the importance of geometric configural constraints in robust feature matching. In the role of geometric constraints in object recognition is studied in depth using edge and other geometric features, and an *interpretation tree* (IT) algorithm is developed to search for globally consistent feature correspondences. In this paper,we present a new method of implementing the geometric configural matching. Compared to the interpretation tree search algorithms whose best-case and worst-case complexities can be significantly different, our method has a very predictable low computational cost and has the best-case and worst-case complexities on the same order.

III. PROPOSED IMAGE REGISTRATION TECHNIQUE

Digital aerial photography acquired with unmanned aerial vehicles (UAVs) has great value for resource management due to flexibility and relatively low cost for image acquisition.[1] The very high resolution imagery (5 cm) allows for mapping bare soil and vegetation types, structure and patterns in great detail. While image acquisition is relatively straightforward, the creation of orthorectified. Image distortion due to the use of low-cost digital cameras, difficulty in locating ground control points and in automatic generation of tie points, and relatively large errors in exterior orientation (camera position and attitude information from the UAV's GPS/IMU).Due to the low payload capability of most small UAVs, imagery is

ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 5, Issue 5, May 2016

often acquired with low-cost, off-the shelf digital cameras. Imagery from those cameras has greater distortion compared to imagery from mapping cameras, and a camera calibration is required to determine the camera's interior orientation parameters.



.
Figure : Steps involved in multimodal image registration

A. Orthorectification

An orthophoto, orthophotograph or orthoimage is an aerial photograph geometrically corrected ("orthorectified") such that the scale is uniform: the photo has the same lack of distortion as a map. Unlike an uncorrected aerial photograph, an orthophotograph can be used to measure true distances, because it is an accurate representation of the Earth's surface, having been adjusted for topographic relief, lens distortion, and camera tilt. [4]

In a perfect world when using imagery to help solve a problem – such as delineating an oil spill, or evaluating crop yields in a precision agriculture project – we would always work with imagery where the information stored within each pixel represents a true location in XYZ, 3-dimensional space. Unfortunately that is not always the case and many times sensors do not deliver product with the precision location needed. Other sensors may generate geolocation information that is not embedded or otherwise associated with the image. In these cases and others, what are some of the options to improve the correlation between information within a pixel and the location it represents in XYZ space?

Georeference:

In the event that an image is delivered without any geolocation information, it is often necessary to utilize an outside source like a Google Maps covering a corresponding area. In a "worst case" scenario it might be necessary to find a feature in the reference source that is also present in the image. Intersections or other non-changing man-made objects make good features to reference. Find the geolocation of that feature from the reference map and attribute the corresponding image pixel location with that information. Your image is now accurately tied to a location on Earth.

Orthorectification:

Images can be orthorectified – which is the process of truly tying a pixel to a real location in 3-dimensional XYZ space – using a mathematical model with rational polynomial coefficients (RPCs) or using a geometric model which considers an internal sensor model. These methods are known respectively as RPC Orthorectification and Rigorous Orthorectification.

Many modern sensors include RPCs with image delivery for just this purpose. For a collection of images that do not have associated RPCs, if you know some key properties about the internal camera orientation and external environment it is often possible to build RPCs. Automated tie point generation, a DEM, and a couple of ground control points can make the process of RPC orthorectification very accurate.

B. Feature Detection:

Feature-based techniques register images by locating image features such as points, lines, and contours and finding correspondences between them. Feature-based matching methods are typically applied when the local structural information is more significant than the information carried by the image intensities.[5]

The scale-invariant region features are used to solve the problems of image registration, suitable and robust similarity measures need to be defined between region intensity values, to deal with multi-modal matching, image noise, and intensity inhomogeneity. Here an entropy-based feature detector is used to select region features that have complex intensity distributions and are stable in both spatial and scale spaces.

This aims to select regions with highest local saliency in both spatial and scale spaces.

For each pixel x on an image, a probability density function (PDF) $p(s, x)$ is computed from the intensities in a circular region any scale described by a radius s centered at x. The local differential entropy of the region is defined by

$$H(s, x) = -\int_R Pi(s, x) \log_2 pi(s, x) di$$

The best scale *Sx* for the region centered at x is selected which has the maximum local entropy. Then the saliency value, $A(Sx, x)$, for the region with the best scale is defined by the extrema entropy value, weighted by the best scale and a differential self similarity measure in the scale space:

$$A(Sx, X) = H(Sx, X) \cdot Sx \cdot \int \| \; pi(s, x) \mid \; sx \| \; di$$

Once the saliency value is calculated for different regions with best scale value, identify the pixels with local maxima in saliency values and pick N most salient ones as region features for the image. The main advantage of this feature is that it is theoretically invariant to scaling, rotation and translation.

C. *Feature matching:*

In the feature matching we are measuring the likelihood of each hypothesized correspondence between a region feature from *If* and a region feature from *Im*, respectively[7]. That is to say, we want to measure the likelihood *Llocal*(*Ci, j*) for each individual feature correspondence hypothesis *C(i, j)* . We can then acquire a total ordering of these hypotheses according to their likelihoods.[5] We define the likelihood to be proportional to the similarity between the interior intensities of the two salient regions involved. Let us denote the *i*th region on *If* as *A*, and the *j*th region on *Im* as *B*. Before measuring their intensity similarity, we first normalize their scales by supersampling (using bicubic interpolation) the smaller region to match the scale of the larger region. This also leads to scale-invariant matching. The translation invariance is intrinsic by aligning the two region centres. To further achieve rotation invariance, we sample the parameter space for rotation sparsely 2, and use the largest similarity value over all possible angles as the similarity between the two regions. The similarity measure we use is a normalized form of mutual information, the Entropy Correlation Coefficient (ECC).Such metric has been proven robust in the literature in dealing with multi-modal image matching, image noise and intensity inhomogeneity.

Formally, the likelihood of a correspondence hypothesis *C(i, j)* is defined as:

$$\pounds_{locall}(C_i, j) = \max_{\theta} ECC(A, B_{\theta})$$

where is the scale-normalized region *B* after rotating angle .
The Entropy Correlation Coefficient (ECC) between
the two regions is defined by:

$$ECC(A, B_{\theta}) = 2 - \frac{2\mathcal{H}(A, B_{\theta})}{\mathcal{H}(A) + \mathcal{H}(B_{\theta})}$$

where *H* indicates the joint or marginal differential entropy of the intensity value random variables of the two regions. Given two inputs *u* and *v*, the value of *ECC*(*u*, *v*) has the following properties: *ECC*(*u*, *v*) is scaled to (0, 1), such that 0 indicates full independence and complete dependence between the two inputs. Furthermore, *ECC*(*u*, *v*) increases almost linearly when the relationship between *u* and *v* varies from full independence to complete dependence which makes it an attractive measure of the likelihood that *u* corresponds to *v*.

D. *RANSAC:*

Rasac (random sanple consensus algorithm was proposed by Fischler and Bolles is a general parameter estimation approach designed to cope with a large proportion of outliers in an input data. Unlike many of common robust estimation techniques and m-estimation and least square that have been adopted by a computer vision community from a statistic literature, ransac was developed within the computer vision community.[6]

RANSAC is a re-sampling technique that generates candidate solutions by using the minimum number observations (data points) required to estimate the underlying model parameters. As pointed out by Fischler and Bolles , unlike conventional sampling techniques that use as much of the data as possible to obtain an initial solution and then proceed to prune outliers, RANSAC uses the smallest set possible and proceeds to enlarge this set with consistent data points RANSAC is a re-sampling technique that generates candidate solutions by using

RANSAC algorithm:

1. Select randomly the minimum number of points required to determine the model parameters.
2. Solve for the parameters of the model.
3. Determine how many points from the set of all points fit with a predefined tolerance
4. If the fraction of the number of inliers over the total number points in the set exceeds a predefined threshold, re-estimate the model parameters using all the identified inliers and terminate.
5. Otherwise, repeat steps 1 through 4 (maximum of N times).

E. *Affine Transformation:*

The affine transformation technique is typically used to correct for geometric distortions or deformations that occur with non-ideal camera angles. the affine transformation is used to match the reference google satellite image with the current UAV image by performing the translation, rotation ,scaling and scaling on the feature matched image to get the

- Translate moves a set of points a fixed distance in x and y.
- Scale scales a set of points up or down in the x and y directions.
- Rotate rotates a set of points about the origin.
- Shear offsets a set of points a distance proportional to their x and y coordinates.

IV. Conclusion

To conclude, we have presented a novel image registration method based on the region component and configural matching using scale-invariant salient region features. The proposed method possesses characteristics of both feature based and intensity-based methods. While the overall framework is based on finding correspondences between features, all the feature correspondence likelihoods and decisions are made according to intensity-based similarity measures between region features and images. The method is efficient in that it recovers a transformation using sparse salient region feature correspondences. It is also very robust because it exploits strict global geometric constraints when finding a joint correspondence between multiple feature pairs.
In our future work, we will extend our algorithm to deal with more complicated transformation models such as affine, projective transformations as well as non-rigid deformations in both 2D and 3D. It is also interesting to investigate schemes to couple the joint correspondence detection and transformation model prediction. The goal is to identify as many good feature correspondences as possible, and fully utilize these correspondences to predict an appropriate transformation model for registration, then to estimate the transformation parameters.

References

[1] Image registration methods: a survey
Barbara Zitova´*, Jan Flusser Academy of Sciences of the Czech Republic
Pod voda´renskou vеˇzˇı´ 4, 182 08 Prague 8, Czech Republic 26 June 2003

[2] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II:264–271, 2003.

[3] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[4] A procedure for orthorectification of sub-decimeter resolution imagery obtained with an unmanned aerial vehicle (uav) Andrea S. Laliberte, Remote Sensing Scientist Craig Winters, GIS Specialist New Mexico State University Jornada Experimental Range, Las Cruces, NM 88003

[5]. Hybrid Image Registration based on Configural Matching of Scale-Invariant Salient Region Features Xiaolei Huang1, Yiyong Sun2, Dimitris Metaxas1, Frank Sauer2, and Chenyang Xu2 1 NJ, USA ƒyiyong. sun

[6] F. Martin and B. Robert, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, Vol. 24, No. 6, pp. 381-395, 1981

[7]A Robust Approach for Multi-Modal Image Registration Nikhila unni1, Preetha V.H
International Journal of Innovative Research in Science, Engineering and Technology , July 2015