POSITION ESTIMATION OF AERIAL VEHICLE BASED ON A VISION AIDED NAVIGATION SYSTEM

F. Samadzadegan [a], M. Hahn [b,*], S. Saeedi [a]

[a] Department of Geomatics Engineeering, University of Tehran, Tehran, Iran – (samadz, ssaeedi)@ut.ac.ir
[b] Dept. of Geomatics, Computer Science and Mathematics, Stuttgart University of Applied Sciences, Stuttgart, Germany - michael.hahn@hft-stuttgart.de

KEY WORDS: Navigation, Pose Estimation, Geo-Referencing, Vision-Based, Matching, Fuzzy reasonning

ABSTRACT:

Automatic pose estimation remains to be one of the challenging problems in photogrammetry and computer vision. When other systems (e.g. GPS and INS) are inadequate, and real time precise positioning is required, image is a very high quality device for localization of vehicles. In this paper, we address a method for Vision Based Navigation (VBN) of Aerial Vehicles which is to determine precise position and attitude of the vehicle in real time. Pose Estimation is achieved by registering single sensed image frames to stored reference database using salient features and descriptors. The reference database includes previously collected satellite imageries and 3D terrain maps that have been precisely aligned to geo-coordinates. The pose estimation strategy has been tested on a number of different images and experimental results proved the feasibility and robustness of our approach.

1. INTRODUCTION

Recently, most of the aerial vehicles use GPS (Global Positioning System) or INS (Inertial Navigation System) for self-localization. But GPS is sensitive to signal dropout and the drawback of INS is that their position error compounds over time and causes large localization errors. In order to overcome main disadvantages of both methods, many researchers suggested the vision-based navigation, which helps estimating pose and orientation of aerial vehicles when GPS or inertial guidance is not available (Gracias et. al., 2003; Naval et. al., 1997; Rushant and Spacek, 1997; Nuno Gracias, 2004).

In this paper we present a fuzzy based 2D-3D pose estimation methodology. With 2D-3D pose estimation we mean to fit 2D sensor data (aerial images) into a 3D object model (ortho-rectified satellite imageries). That is we are interested in estimating a rigorous transformation with minimizes a similarity measure between image and object data (Figure 1). This approach develops an automatic system for simultaneously precise determining the position and attitude of the aerial vehicle as well as the correspondence between image and object features.

The body coordinate system (*body*-System) used in navigation applications seems to be similar to the camera coordinate system (*Cam*-System) used in geo-referencing. After camera calibration process (Interior Orientation) the camera coordinate system is realized by the fiducial marks of the camera or the CCD sensor. The origin is the projection centre *P* in the distance of the focus length *f* to the principal point (Figure 3).
Instead of the navigation system (*Nav*-system) in geo-referencing the quite similar earth fixed terrain or object coordinate system (*Obj*-System) is used. *Obj*-System established by the national map projection, spanned by the grid east-, north-, and up-axes in the UTM projection and WGS 84 Ellipsoid. The transformations between *Cam*-system and *obj*-system are established by Co-linearity condition.

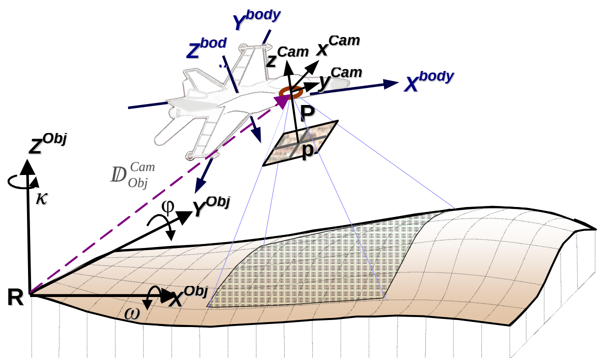


Figure 1. Definition of the coordinate system and vectors used for Geo-referencing

2. PROPOSED METHODOLOGY

In the proposed pose estimation methodology, the aerial images have to be registered by geo-referenced image database. The overall strategy for proposed pose estimation approach may be expressed in two distinct phases: 1. Create geo-referenced database before mission, and 2. Automatic image geo-referencing and pose estimation of Vehicle during the mission. In the following, the main components of the each phase will be described with more details (see figure 2).

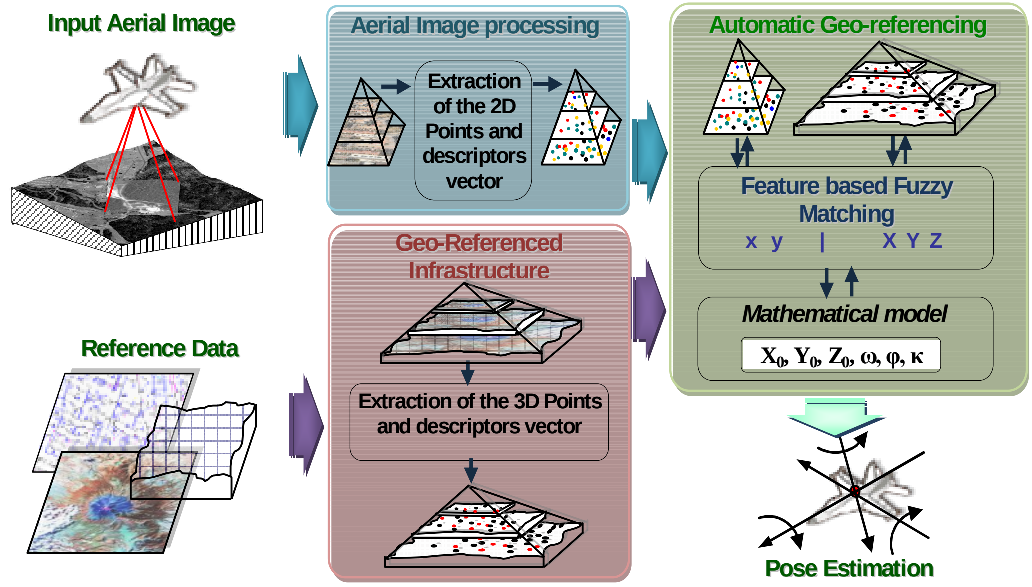


Figure 2. Propose automatic pose estimation approach

2.1. Create Geo-Referenced Database

The ortho-rectified images are an important layer of topographic information that is generated from satellite Imageries and corresponding DSM (Digital Surface Model) by rectification process (Figure 3). It can be serves as 3D object space.

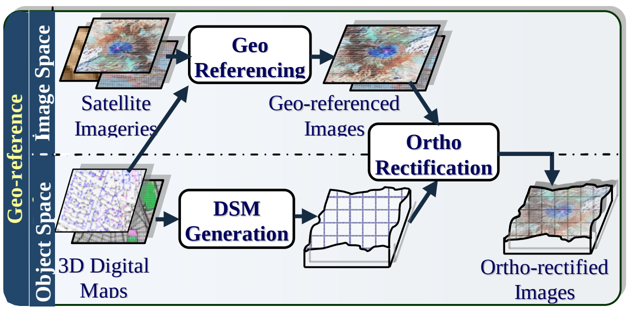


Figure 3. Propose automatic pose estimation approach

2.1.1 Feature and descriptor Extraction: This step is based on the extraction of salient features and descriptors vector in the 3D object space (ortho-rectified Image). Significant points (salient point, region corners, line intersections, etc.) are understood as features here that are distinct, spread all over the image and efficiently detectable in both spaces.

The automatic selection of image points is based on the well-known operator proposed by Harris and Stephens (1988). The Harris operator is a popular interest point detector due to its strong invariance to: rotation, scale, illumination variation and image noise (Schmid et. al., 2000). The Harris corner detector is based on the local auto-correlation function of an image; where the local auto-correlation function measures the local changes of the image with patches shifted by a small amount in different directions (Harris 1987). Given a shift (Δx, Δy) and a point (x, y), the auto-correlation function is defined as:

$$c(x,y)=[\Delta x \quad \Delta y]\cdot\begin{bmatrix}\sum_W (I_x(x_i,y_i))^2 & \sum_W I_x(x_i,y_i)I_y(x_i,y_i)\\ \sum_W I_x(x_i,y_i)I_y(x_i,y_i) & \sum_W (I_y(x_i,y_i))^2\end{bmatrix}\cdot\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} \quad (1)$$

where $I_x(.,.)$ and $I_y(.,.)$ are Partial derivatives of image function in x and y, (x_i, y_i) is The points in the window W *and* W is Gaussian filter centered on (x, y). The summation is performed over a predefined neighbourhood area. Finally a non maximum suppression is performed and threshold value determines the final points.

For each extract features, descriptor vectors which include textural and spectral information derived from pictorial patch surrounding the extracted points, generated. The descriptors vectors are highly robust and distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from reference infrastructure (Figure 4).

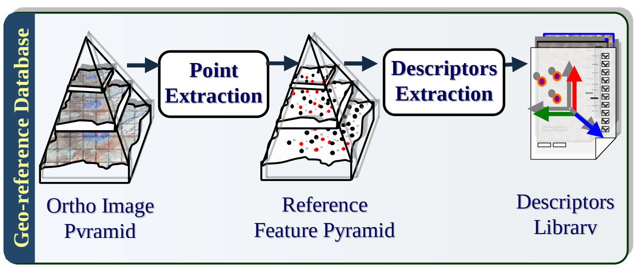


Figure 4. Propose automatic pose estimation approach

Before the mission, the derived coordinates and descriptors vectors of feature will be generated in different resolution pyramids and restored by additional layers in the geo-reference database.

2.2. Automatic Image Geo-referencing

In this study the problem of automatically registration of the sensed aerial images and a 3D model (ortho-rectified image) of the corresponding scene using a priori information is considered. Due to the diversity of image acquisition (different viewpoints, different times, and different sensors) and due to various types of degradations (light condition, occlusion, shadow, relief), registration procedure encounters geometric and radiometric deformation. Proposed method should take into account not only the assumed type of geometric deformation between the images but also radiometric deformations and noise corruption, required registration accuracy and data characteristics.

The proposed methodology is accomplished in a coarse-to-fine approach with three levels of coarse, intermediate and fine resolution to achieve reliable, stable and precise registration respectively. This can provide rough approximate values for the successive levels of image pyramids. Construction of image pyramids in this project is carried out according to wavelet transform. The wavelet transform features are used because wavelet transforms convey both space and time characteristics and their multi-resolution representations enable efficient hierarchical searching (Daubechies, 1992; Dianat and Rao, 1998).

In each level according to tree basic steps of registration procedure we will deal with the feature extraction, similarity measures and matching strategy. In this step during the mission, the multi resolution representation of aerial image must be constructed and then in each layer of pyramid, salient features and relevant descriptors will be extracted. Simultaneously based on a fuzzy reasoning system, conjugate points will be determined by using descriptors vector.

2.2.1 Fuzzy Based Image Matching: Although in the classical image matching approaches radiometric and geometric conditions are effective tools for the determination of the conjugate points, in practice, however, the main problem with these methods is their inability to reliably and realistically fuse these conditions for decision making. There are several schemes for the fusion of different conditions (Samadzadegan, 2002). Fuzzy reasoning is one of these methods by which the parameters that influence the decision making process can be realistically fused using a human like reasoning strategy. This is achieved by defining the so called linguistic variables; linguistic labels and membership functions (Zadeh, 1965; Stoms, 1987). The fuzzy reasoning process is then realized using the fuzzy if-then rules that enable the linguistic statements to be treated mathematically (Zadeh, 1965; Stoms, 1987). The matching success depends on how well the linguistic variables, membership functions and the related rules are defined.

Our proposed linguistic variables (Table 1) are *Geometric* (*y*-discrepancy and *x*-discrepancy) and *Radiometric* (correlation coefficient and texture rank differences).

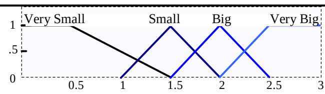
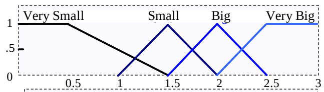
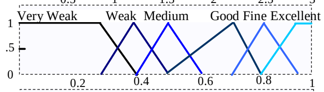
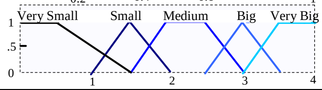
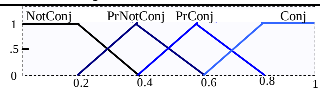
Geometric variables: The imaging geometry is defined by the exterior orientation parameters of the imaging system. The geometric parameters are considered to include the imaging geometry represented by *y*-discrepancy and also *x*-discrepancy between transformed extracted 3D pointes by co-linearity condition and extracted 2D points in image space.

Radiometric variables: The semantic conditions are defined based on the radiometric similarities between the conjugate points. This can be determined via different similarity assessment algorithms. Our method takes advantage of two different algorithms, namely: the well known normalized correlation coefficient (NCC) and the rank differences. The correlation function between two images A_{NxN} and B_{NxN} is given as:

$$NCC(A,B) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}((A_{ij}-\bar{A})(B_{ij}-\bar{B}))}{\sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(A_{ij}-\bar{A})^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(B_{ij}-\bar{B})^2}} \qquad (2)$$

The so called rank values are computed using window arrays constructed around already generated salient points. The rank of a window is an integer value denoting the gray shade rank of the central pixel as compared with other pixels of the window. It is assumed that conjugate features should demonstrate a rather similar rank values. Our experiment with different data sets show that the rank values can contribute effectively to the determination of the conjugate features.

Table 1. Linguistic Variables, labels and Membership function for the fuzzy based image matching

Variables		Linguistic Variables	Linguistic Labels	Membership function and the corresponding labels
Input	Geometric	ΔY	Very Small, Small, Big, Very Big	
		ΔX	Very Small, Small, Big, Very Big	
	Radiometric	Correlation	Very Weak, Weak, Medium, Good, Fine, Excellent	
		ΔRank	Very Small, Small, Medium, Big, Very Big	
Output	Conjugate	Conjugate	Not Con, Probably Not Con, Probably Con, Con	

For each of the linguistic variables, membership functions are defined by an experienced operator. The membership function for the geometric constraint (i.e. *x*- and *y*-discrepancy) is determined based on the geometry, scale and resolution of the image. The membership function and the labels for *y*-discrepancy variations are presented in table 1.

2.2.2 Fuzzy Rules: Having determined the linguistic variables and their corresponding membership functions, it is now possible to determine the fuzzy rules between the input and the output membership functions. These rules are determined using an experienced human operator. Based on the defined membership functions and the rules, fuzzy reasoning for the conjugate point determination is carried out in a Mamdani type fuzzy reasoning structure (Zadeh, 1965). After defuzzification process, the candidate key points are categorized as matched points. In the following some sample rules for conjugate point determination are presented (Samadzadegan, 2002):

- **IF** *discrepancy X* Is *Very Small* **AND** *discrepancy Y* Is *Very Small* **AND** *Correlation* Is *Fine* **AND** *Rank Difference* Is *Small* **THEN** *Key Points* Are *Conjugate,*
- **IF** *discrepancy X* Is *Medium* **AND** *discrepancy Y* Is *Small* **AND** *Correlation* Is *Good* **AND** *Rank Difference* Is *Small* **THEN** *Key Points* Are *Probably Conjugate,*
- **IF** *discrepancy X* Is *Medium* **AND** *discrepancy Y* Is *Small* **AND** *Correlation* Is *Medium* **AND** *Rank Difference* Is *Small* **THEN** *Key Points* Are *Probably Conjugate,*
- **IF** *discrepancy X* Is *Medium* **AND** *discrepancy Y* Is *Medium* **AND** *Correlation* Is *Good* **THEN** *Key Points* Are *Probably Not Conjugate,*

3. EXPERIMENTAL RESULTS

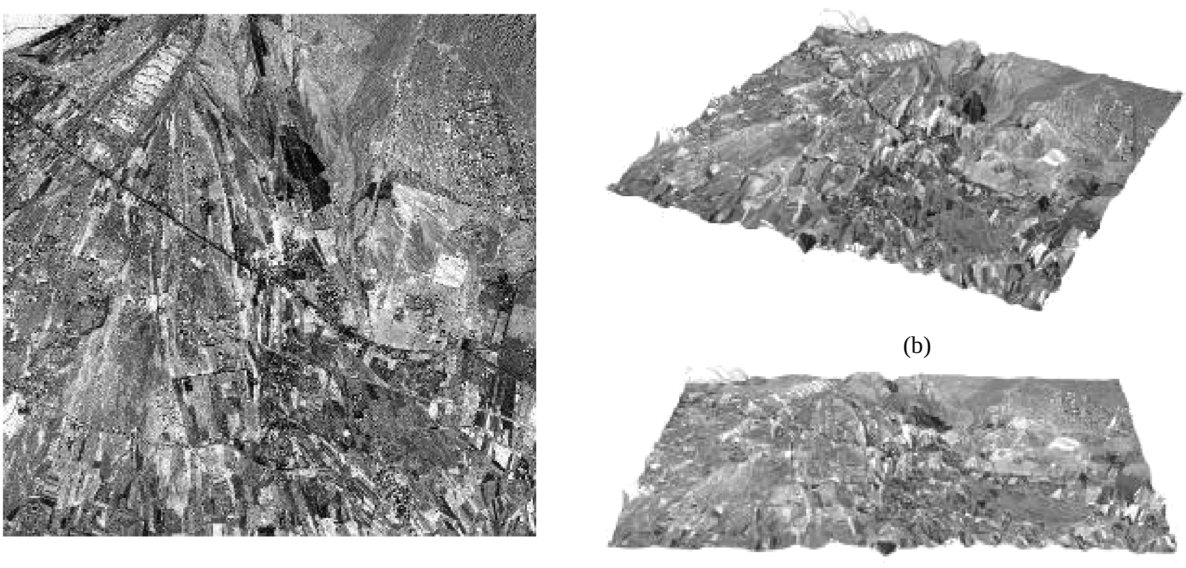
The potential of the proposed was evaluated through comprehensive experimental tests conducted on a wide variety of datasets. In this paper the tests are conducted on a geo database consist of the Geo panchromatic IKONOS imageries with ground resolution of about 1 meter. The test image was taken over the city of Sharifabad, Iran on 17th of December, 2005 (Figure 5). The relief variation in the area is in the range of 1500m to 1900m above sea level (ASL). The aerial vehicle was flown at an altitude of about 1500 meters above ground resulting in an image scale of 1:10000, respectively. The focal length of aerial camera was 153.86 mm and CCD chip is 1024x1024 pixels with 9µm pixel size.

The proposed approach categorize by two phases: Pre-mission phase and during-mission phase.

Pre-mission phase: The rectification process is carried out before mission with PCI Geomatica software and the ortho-rectified images store in a spatial data base with three different resolution levels (Coarse, intermediate and fine level). Then Feature pyramids and their descriptor vectors are stored based on derived point coordinates. To Ortho-rectify the satellite imagery, the correspondence DSM was produced from two different scale maps: 1:2000 and 1:25000. these two maps produce two different DSM, that the first one has 1 m resolution with .2 m vertical accuracy and the other on has 10 m resolution and 2.5 m vertical accuracy. Figure 7 shows the IKONOS image and 3D view of the two different ortho-rectified Images.

During-mission phase: In this stage, after 2D point extraction of sensed aerial image, a hierarchical (coarse-to-fine) feature based image matching is performed to determine the conjugate points. In each level, the correspondence between the features detected in the sensed image and those detected in the reference ortho-rectified image is established based on a fuzzy reasoning system.

Table 2 shows the independent results for each pyramid layer obtained by the fuzzy image matching strategy. A comparison between the number of the detected features in each layer and the number of matched points clearly indicates how the proposed strategy has eliminated some of the points in each layer (see Table 2). These are the points for which, the geometric and the radiometric conditions have not been satisfied according to the fuzzy reasoning system. The RMSE values obtained by the fuzzy based registration method are also given in Table 2.



(a) (b) (c)

Figure 5. Reference geo database used for the evaluation of the proposed method. a) IKONOS panchromatic image, b) the corresponding 3D view of the reference orthophoto witch is produced by 1:2000 scale map and c) the corresponding 3D view of the reference orthophoto witch is produced by 1:25000 scale map.

Table2. Results for each pyramid layer obtained by the Fuzzy image matching strategy

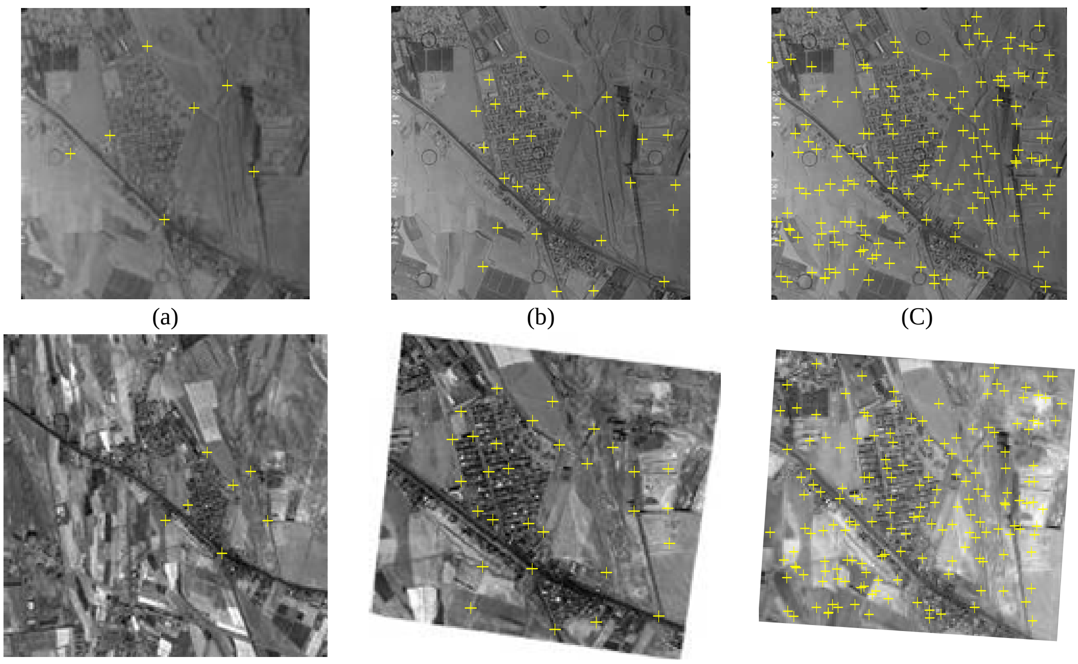
Referenced DSM	***Layer***	***Left Image Points***	***Right Image Points***	***Matched Points***	***RMSE (m)***		***Max Error (m)***	
					X	**Y**	**X**	**Y**
1:25000	**Coarse**	*12*	*18*	*7*	*10.40*	*10.56*	*11.07*	*12.56*
	mediate	*37*	*42*	*28*	*5.08*	*6.02*	*6.73*	*6.21*
	Fine	*109*	*102*	*86*	*2.04*	*1.82*	*2.83*	*2.96*
1:2000	**Coarse**	*12*	*18*	*7*	*9.22*	*10.56*	*11.05*	*11.96*
	mediate	*37*	*42*	*28*	*6.78*	*6.28*	*7.21*	*7.89*
	Fine	*109*	*102*	*86*	*2.02*	*2.85*	*3.59*	*3.06*

In the *coarse* level, in order to estimate the reliable corresponding point, we used fuzzy matching algorithm by the initial value of mapping transformation (provided by auxiliary data). The next level in the pyramid (*intermediate*) is then used to achieving stable strategy. At this level the estimated parameters of mapping transformation from *coarse* level are further refined via matching algorithm. Finally the *fine* resolution level in the pyramid is then processed. The initial estimated parameters of mapping transformation at this level are again refined via new conjugate points to output the final accurate and precise result of exterior orientation parameters. In this geo-referencing process, while looking for the feature correspondence, simultaneously update the estimation of the mapping functions parameters and thus merge the second and third Geo-referencing steps (Figure 6).

Next we deal with the issue of determination the 3D position and orientation of the vehicle. The problem of pose estimation is equipment using the available information on the geo-referencing parameters. Table 3 shows the results for each pyramid layer obtained by our approach. The calculated parameters especially when compared to those obtained through manual procedures clearly indicate the ability of the purposed methodology.

4. CONCLUSIONS

In this paper we presented a new approach for vision-based pose estimation of unmanned aerial vehicles that affords reliable estimates of the aircraft position and orientation. In the proposed methodology, we present an algorithm that robustly aligns an Aerial Image to an Area Reference Image while realistically updating the sensor model parameters. Our method uses feature based registration of aerial and satellite imageries based on a fuzzy reasoning system. These feature based method is invariant with respect to translation, rotation, and scale (up to a factor of four). Furthermore, the system uses local features and therefore is robust to changes in the scene. The implemented methodology has proved to be very efficient and reliable for automatic aerial vehicle pose estimation. This approach is particularly useful in the context of aerial mobile robots and UAVs.



(a) (b) (C)

(d) (e) (f)

Figure 6. Tree layer of multi resolution representation of Aerial sensed image and corresponding satellite imagery with matched points: a)Layer 2 of Arial Image pyramid with matched point, b)Layer 1 of Arial Image pyramid with matched point, C)Layer 0 of Arial Image pyramid with matched point, d)Search area of satellite image with matched point (Layer 2), e)Search area of satellite image with matched point (Layer 1), and f)Search area of satellite image with matched point (Layer 0).

Table3. The results for each pyramid layer, initial value and manual approach by control points

Ref. DSM	*Description*		*X (m)*	*Y (m)*	*Z (m)*	*ω°*	*φ°*	*κ°*
-	**Initial values**		572072	3920098	2302	0	0	0
-	**Manual approach**		572024.0334	3920021.8096	2331.1128	-0.4218	-0.1983	2.6920
1:2000	**Proposed approach**	**Coarse**	572029.5454	3920027.3216	2336.6248	0.4924	0.7159	3.6062
		mediate	572023.4268	3920021.2030	2330.5063	-0.2 741	0.3494	2.9397
		Fine	572023.7981	3920021.9874	2330.5413	-0.444599	-0.214711	2.708699
1:25000	**Proposed approach**	**Coarse**	572014.3050	3920012.0812	2321.3845	1.3956	1.6191	4.5094
		mediate	572020.2320	3920018.0082	2327.3114	0.7948	0.5713	3.6685
		Fine	572023.2262	3920021.0023	2330.3056	-0.5547	-0.7782	**3.1385**

References

Harris C., 1987. Determination of ego-motion from matched points. In proceeding *Alvey Vision Conference*, Cambridge, UK.

Harris C. and M.J. Stephens, 1988. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–152.

Schmid C., R. Mohr, and C. Bauckhage, June 2000. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.

ARINC 705, 2001. *Airlines Electronic Engineering Committe (1982)*, ARINC Characteristic 705 - Attitude and Heading Reference System. Aeronautical Radio INC., Annapolis, Maryland. http://www.arinc.com/cgi-bin/store/arinc (May 6, 2004).

Bäumker, M., 1995. Basiswissen Inertial- und Sensortechnik. In: Journal for Satellite-Based Positioning, *Navigation and Communication*, Vol. 4/95, pp. 147-152, December 1995.

Daubechies, I, 1992. Ten Lectures on Wavelets, *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM.

Dianat S. A. and R. Rao, 1998. “Wavelet Transforms: Theory and Applications,” *SPIE’s 12th Annual International Symposium on Aerosense/Defense Sensing, Simulation, & Controls*, 13-17 April.

Nuno Gracias, Sjoerd van der Zwaan, Alexandre Bernardino, José Santos-Victor, 2003. Mosaic Based Navigation for Autonomous Underwater Vehicles, VisLab-TR 06/2003, *IEEE Journal of Oceanic Engineering*, October.

Naval P. C. Jr., M. Mukunoki, M. Minoh, and K. Ikeda, 1997, “Estimating Camera Position and Orientation from Geographical Map and Mountain Image.” *38th Research Meeting of the Pattern Sensing Group, Society of Instrument and Control Engineers*, pp.9--16.

Rushant K., L. Spacek, 1997. "An Autonomous Vehicle Navigation System using Panoramic Machine Vision Techniques" Technical report CSM-298.

Samadzadegan, F., 2002, Automatic 3D Object Recognition and Reconstruction Based on Artificial Intelligence and Information Fusion Techniques. Ph.D. thesis, Tehran University, Faculty of Engineering, Dept. of Geomatics Engineering.

Stoms, D., 1987. Reasoning with uncertainty in intelligent geographic information systems, Proceeding, *GIS 87*, 693-700

Zadeh, L. A., 1965. *Fuzzy Sets, Information and Control*. Vol. 8, pp. 338-353.