International Journal of Computer Vision and Signal Processing, 2(1), 29-37(2013) ORIGINAL ARTICLE

Choosing Appropriate Homography Transformation for Building Panoramic Images

Prakash Duraisamy*, Yassine Belkhouche

Department of Computer Science
University of North Texas
Denton-76203, Texas, USA

Stephen Jackson**

Department of Mathematics
University of North Texas
Denton-76203, Alabama, USA

Kamesh Namuduri

Department of Electrical Engineerng
University of North Texas
Denton-76203, Texas, USA

Bill Buckles

Department of Computer Science
University of North Texas
Denton-76203, Texas, USA

Abstract

In building panoramic images, the selection of appropriate homography plays a crucial role in reducing the error and in registering the images accurately. In this paper, we demonstrate a method for selecting the appropriate homography for building the panoramic image based on information extracted from the images. It is shown that using homographies from the appropriate subgroup, the undesirable distortions can be reduced which improves the quality of the panoramic image. We tested our method both on synthetic and real world images. We also discussed and compared several error metrics to evaluate the accuracy of registration.

IJCVSP
International Journal of Computer Vision and Signal Processing

ISSN: 2186-1390 (Online)
http://www.ijcvsp.com

Article History:
Received: 1 July 2012
Revised: 1 September 2012
Accepted: 20 September 2012
Published Online: 25 September 2012

Keywords: image registration, projective transformation, homography,

1. INTRODUCTION

Image registration is a process of matching images based on a geometric transformation of one or more images. It is a prior step before image fusion. The quality of the final output depends upon how well the images are aligned. Image mosaicing is the technique which aligns the multiple images into one

*Corresponding author
**Principal corresponding author
Email addresses: prakashduraisamy@my.unt.edu (Prakash Duraisamy), myb0012@unt.edu (Yassine Belkhouche), jackson@unt.edu (Stephen Jackson), kamesh.namuduri@unt.edu (Kamesh Namuduri), bbuckles@cse.unt.edu (Bill Buckles)

panoramic image which gives an extended field of view. This technique is used in many applications such as creating terrain maps from high altitude images. Image mosaicing has many more new applications including map updation, virtual reality, and medical imaging. In general, the underlying principle of 2D image mosaicing is that of the homography. A homography is an invertible mapping between two images [13]. It is a linear transformation when the images coordinates are viewed as being in projective 2-space (so homography transformation H is a 3×3 matrix). A homography is not only responsible for building panoramic images, but also its accuracy will decide the final resolution of the image. Minimal errors in the homography will lead to a seamless mosaic image.

The homography is computed based on an overlapping part between the two images (the source and reference images). The accuracy of the homography depends largely upon how well feature points are extracted and matched, but in reality there are several limitations. Several techniques such as SIFT (Scale Invariant Feature Transform) and RANSAC (Random Sample Consensus) eare commonly used to improve the point matching process. After matching feature points, homography is computed mathematically from the matched point data sets. There are several methods for computing the homography, such as the Newton-Raphson algorithm, Levenberg Marquardt algorithm, and the Sampson approximation algorithm. However, the quality of the output of homography cannot exceed theoretical boundaries which depend on the errors in the matched data points and errors in the images themselves (the images may not be truly planar). Also, in some situations, feature points in one image (source image) may lie on or beyond the frame in the other image (reference image). This may be due to some feature points lying on points at infinity which causes instability in the mapping. An algorithm was introduced for accurate mapping using a technique called renormalization [6]. Although the renormalization technique improves the mapping to certain extent, it does not remove the distortion completely.

In the past, most of the work focussed on improving the post estimation of the homography. In this paper, we demonstrate that improvements can be made if an appropriate subgroup of homographies is considered in the homography computation, and we give a method (not involving a priori scene information) for selecting this subgroup. The subgroup selection can be thought of as extra scene information, but we wish to have this determination made by the algorithm itself. The subgroups of homographies (see Fig 1) which are relevant to our method are the full group of **projective homographies** (8 degrees of freedom in the matrix), their sub-groups of the **affine transformations** (6 degree of freedom), the **similarity transformations** (4 degrees of freedom), and the **Isometry trasformations** (2 degrees of freedom). Other subgroups could certainly be considered as well. This approach of pre-selection of the homography type will reduce the error and improve the final resolution of the image. It is perhaps not obvious at first glance that this should be the case, as all of the above mentioned homography types are certainly projective homographies, and one might think at first that treating them as such would give the best possible estimates for the homography (perhaps at the expense of some computation time).

Early registration techniques were mainly intended for creating a seamless mosaic between overlapping im- ages [4], [7]. Several mosaicing methods are reviewed in [5]. High quality panoramic images can be created if the camera parameters are known. However, complex methods are used to join the overlapping images when the camera parameters are not available. In order to improve the homography produced by the Discrete Linear Transform (DLT) algorithm, many strategies such as the Levenberg-Marquart (LM) algorithm have been developed. In [10], the LM algorithm is shown to be better than

the DLT algorithm when joining sequential images. There are other optimization techniques such as the Sampson approximation [13] which also improve the homography. Even though these methods improve the homography, they leave open the possibility of improvement as they do not make use of prior knowledge about the appropriate homography sub-class model. In this paper, we demonstrate that selection of the appropriate subgroup type will in fact lead to more accurate registration. Thus, we can improve the quality of the final mosaic image by incorporating subgroup selection in the process. Our objective is to investigate the criteria for subgroup selection ("model selection"), and show that this criteria can be obtained from the image data itself.

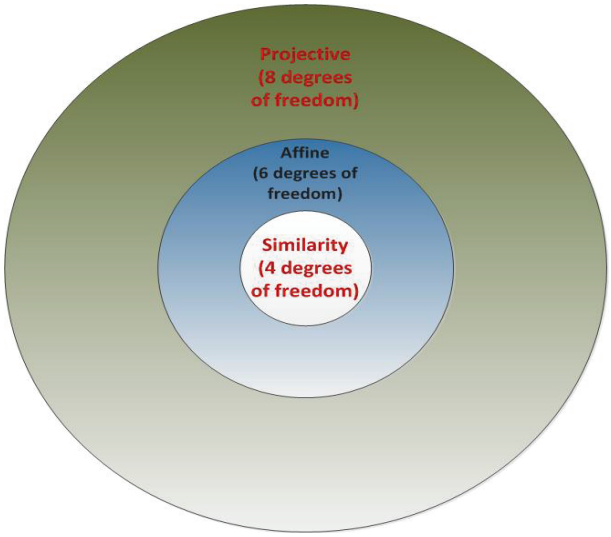


Figure 1: Hierarchy of Transformations

1.1. Data Acquisition

We tested our algorithm on two kinds of data sets: sythetic and real-world. The synthetic data was generated using Maple CAS and was analyzed entirely without involving any manual process. For the real-world data, we have taken a portion of the New Orleans area captured using a digital camera (with fixed focal length) from a helicopter, and of a scene in the Denton, Texas area captured by an aerial fly-over. Some of the images involved a significant oblique perspective, i.e., they involve significant effects of non-planarity, so as to test the limitations of the method.

2. Related work

Image registration methods are divided into two categories. Feature based techniques are considered when local structures are dominant as opposed to image intensities. In most cases, feature based techniques are preferred over direct correlation methods due to their more robust nature. In the past, registration is limited to images that have undergone simple translations. Later,Fast Fourier Transforms and cross correlation techniques were introduced in [1] to determine spatial distances between the images. In a different approach [12], edge correlation is introduced. This technique assumes that edges are more invariant to image effects than other features like texture, color and intensity. An enhanced version of [12] was illustrated in [2]. In this approach, phase correlation which gives rotation as an additional parameter, is included. For joining images, the phase component is used in [8]. This work is extended in [11] to perform registration based on rotation, translation and scaling differences.

All the above techniques, compute the homography based on their individual approach which will not be suitable for other subgroups of homographies. In addition, the above techniques compute the error after estimating the homography. Later, a new technique for improving the homography based on map registration is introduced in [9]. But, this technique still fails to work on all subgroups of homographies. In contrast, the above problem was solved in [6] by introducing the notion of stabilizing the mosaicing by model selection. Even though this model solves the post homography error problem using geometric AIC, it not robust in all situations. In our approach, in addition to above model, we introduce subjective and numerical analysis which

will offset the problems of above techniques.

In [3], we demonstrated that the appropriate selection of the homography type could reduce the registration error, but methods for homography model selrction were not presented. Here we attempt to demonstrate that methods based on the images alone, with no extra scene or camers information, can be used to guide the subgroup selection.

3. HOMOGRAPHIES

In this section, we discuss basic facts about homographies and their subgroups briefly before presenting our method. In our approach, we are not assuming any apriori knowledge about the intrinsic parameters of the camera, or of the nature of the homography. Recall that a homography is a geometric transformation between two images such that three points x_1, x_2, x_3 lie on the same line if only if $h(\mathbf{x_1})$, $h(\mathbf{x_2})$, $h(\mathbf{x_3})$ do [13]. Such a map h can be represented as an invertible mapping from projective 2-space (the projective plane) $\mathbf{P^2}$ to itself. It is thus expressed as a non-singular 3×3 matrix H, which we identify with the homography. So, we have,

$$x' = \mathbf{H} \times x, \tag{1}$$

where x and x' are the coordinates of a point in the two images viewed in projective space. The matrix H is only determied up to a scale factor, and thus has 8 degrees of freedom.

We created two synthetic images, namely points x (first image) and x' (transformed second image), each consisting of 20 coordinate points, in the Maple environment. Let (x_i, y_i) (for $1 \leq i \leq 20$) be the Euclidean point coordinates in the first synthetic image and (x'_i, y'_i) be the two coordinates in the second image. We start our analysis with the random generation of the points (x_i, y_i) and a randomly generated homography H, which will serve to generate the theoretical "true homographies." Let I_1 denote the first image consisting of the 20 points (x_i, y_i). We generate 21 theoretical homographies H_t, ranging from purely affine to general projective, by fixing all the entries of H except for the last row (that is, $H_t = H$ except for the last row), fixing $H_t(3,3) = 1$, and letting the entries $H_t(3,1)$, $H_t(3,2)$ vary linearly with t from values 0 to $H(3,1)$ and $H(3,2)$ respectively. We thus have a sequence of homographies H_t, $0 \leq t \leq 20$, which are closer to affine with smaller values of t (recall that an affine transformation is represented by an H with $H(3,1) = H(3,2) = 0$). Using each H_t, we compute corresponding second images I_2^t by $I_2^t = H_t I_1$. We next introduce random Gaussian noise into both the first image and the set of second images (we used the same noise for all of the second images, and a different noise for the first image). Let us call the noisy images N_1, N_2^t. Starting from the noisy images only, we then compute a homography which best matches these images. This is done in two different ways: by assuming the homography lies in the subgroup of affine homographies, and by treating it as a general projective map. Thus, for each value of the parameter t we have an affine homography H_t^a, computed according to the assumption that H is an affine mapping, and a homography H_t^p computed with no assumptions on H.

To be specific, H_t^a, H_t^p are computed as follows. For the projective case, following [13] chapter 4, we consider the equations $H\boldsymbol{x_i} \times \boldsymbol{x}'_i = \mathbf{0}$, where $\boldsymbol{x}_i = \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix}$ are the coordinates of the ith point from the first image in projective coordinates, and likewise for $\boldsymbol{x}'_i$. Each pair of matched points $(\boldsymbol{x_i}, \boldsymbol{x}'_i)$ gives 3 homogeneous equations with variables the entries of H. however, only two of the three equations are independent. So, we get a $2n \times 9$ system, where $n = 20$ is the number of points in the images. An optimal solution to this system in the least squares sense is computed from the singular values decomposition (SVD) of the $2n \times 9$ coefficient matrix (the singular vector corresponding to the smallest singular value). For the affine

case, we follow the same procedure, but we require that the variables corresponding to $H(3,1)$ and $H(3,2)$ are set equal to 0, that is, we use only 7 variables to represent H. So, we get a $2n \times 7$ coefficient matrix for the system, for which we find an optimal solution exactly as before.

4. Appropraite Model Selection

In this section, we discuss different ways of identifying the appropriate homography model for building the panoramic image. The proper selection of the homography model plays an important role in reducing the error in the panoramic image. We compare affine and projective model selection, though the methods should generalize to other subgroups of homographies.

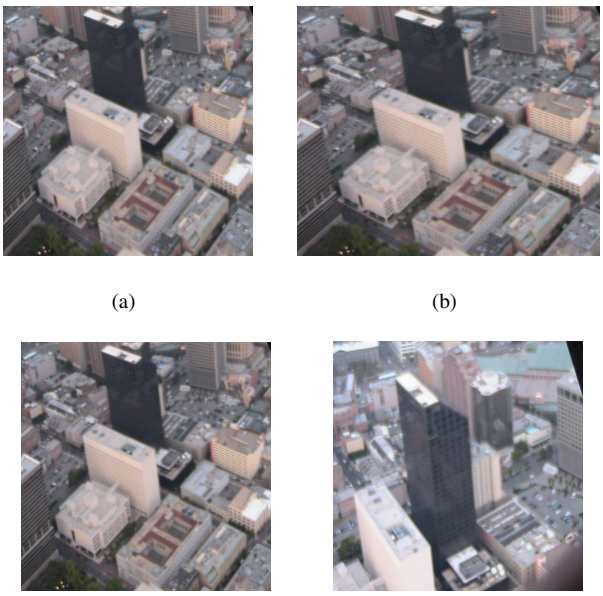
4.0.1. Model Selection using Visual Confirmation

The appropriate model can in some instances be determined visually, up to a certain extent. For example, if the two images only differ by translation, rotations, and uniform scaling in both directions, then it is considered as a similarity transformation. If the scales are different for the two axes, we could consider the affine model. If there is a clear change of perspective, then the general projective model could be used. Examples of these kinds of transformations are shown in Figure 2.

There are cases, such as high altitude orthonormal views, for which these may be considered safe assumptions. However, using visual selection, it is not clear where the crossover point in the model selection occurs, nor does this automate the procedure. Our purpose is to investigate these points.

4.0.2. Model Selection using Homography as a Clue

As we discussed in section III, from the computed homographies H_0, ..., H_{20}, we can classify H into the appropriate subgroup of homographies. If the values of h_7 and h_8 in the last row of the computed homographies $H_1, \ldots, H_{20}$ are zero, then the homography is considered as purely affine. If h_7 and h_8 are far from zero then it is considered as a projective transformation. If the numerical values of h_7 and h_8 are close to zero, then the transformation would naturally be considered as close to affine, but still it is technically considered as projective. In contrast, if the numerical values of h_7 and h_8 are close to 1 or larger, then it surely would best be considered a general projective transformation. So, from the numerical values of h_7 and h_8 we could base our model selection. The exact cutoff point will be investigated below.



(a) (b) (c) (d)

Figure 2: (a) Original, (b) Similarity, (c) Affine, (d) Projective

The homography matrices $(H_0, \ldots, H_{20})$ corresponding to the 21 samples are given below. The h_7 and h_8 values confirm the above discussion, and the numerical values of h_7 and h_8 show that image is moving gradually from affine to projective.

$$H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ \mathbf{h_7} & \mathbf{h_8} & h_9 \end{pmatrix}$$

Homography Output for 20 samples

$$H_0 = \begin{pmatrix} 0.70179871 & 0.13992647 & 0.099520405 \\ 0.97914866 & 0.23963466 & 0.865872898 \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 0.70179871 & 0.13992647 & 0.099520405 \\ 0.97914866 & 0.23963466 & 0.865872898 \\ \mathbf{0.00109777} & \mathbf{0.03109407} & 1 \end{pmatrix}$$

$$\vdots$$

$$H_{10} = \begin{pmatrix} 0.70179871 & 0.13992647380 & 0.099520405 \\ 0.97914866 & 0.23963466 & 0.865872898 \\ \mathbf{0.01097770} & \mathbf{0.03109407} & 1 \end{pmatrix}$$

$$\vdots$$

$$H_{19} = \begin{pmatrix} 0.70179871 & 0.13992647 & 0.099520405 \\ 0.97914866 & 0.23963466 & 0.865872898 \\ \mathbf{0.02085767} & \mathbf{0.59078744} & 1 \end{pmatrix}$$

$$H_{20} = \begin{pmatrix} 0.70179871 & 0.13992647 & 0.099520405 \\ 0.97914866 & 0.23963466 & 0.865872898 \\ \mathbf{0.02195544} & \mathbf{0.62188151} & 1 \end{pmatrix}$$

4.0.3. Model Selection using Error as Clue

It is known [13] that when H is purely affine, the algebraic error and geometric errors are equal (see [13] for the definitions of these errors). As the transformations move away from affine, we would expect that the difference between the algebraic and projective erors to increase. Thus, we could use the difference $\delta = |E_{\text{alg}} - E_{\text{geom}}|$ to guide the model selection, where E_{alg}, E_{geom} are the algebraic and geometric errors respectively. Since these errors are commonly computed anyway, this could be a cost-free parameter for model selection. Again, we investigate the selection criteria below.

5. SELECTION of APPROPRIATE ERROR METRIC

The computation of image mosaicing using appropriate homographies will lead to seamless and accurate mosaicing. However, homography selection is a preprocessing step for creating panoramic image. Once the homography is computed, the selection of appropriate error metric is also important during post processing. In the following, we describe four different metrics for computing the error after image registration. In our work, we used geometric error metric due to its stability and invariance under Euclidean transformation.

5.1. Mean Square Error

Mean square error (MSE) is a function of pixel intensity difference on the common overlap area of the two images. Let I_1' and I_2' be the warped I_1 and I_2 images and let $A(i,j)$ represent their average value, then the mean square error (MSE) metric is defined as:

$$MSE = \sum_{i,j \in I_1' \cap I_2'} (A(i,j) - I_1'(i,j))^2 + (A(i,j) - I_2'(i,j))^2 \tag{2}$$

where $I1'(i,j)$ represents the intensity value at the point (i,j) of the image $I1'$, and $I2'(i,j)$ represents the intensity value at the point (i,j) of the image $I2'$.

This is a straight-forward metric and is sensitive to image brightness.

5.2. Algebraic Error

To compute the algebraic error (AE) one first computes a set of point correspondences between the two images. This is done using the SIFT and RANSAC algorithms. The point correspondences are denoted as $x_i \leftrightarrow x_i'$. Let H denotes the homography which is joining two images. Each point correspondence will

generate a partial error vector $\varepsilon_i = ||x_i' \times Hx_i||$. AE is the sum of these partial errors [4]:

$$AE = ||\varepsilon||^2 = ||x_i' \times Hx_i||^2 \quad (3)$$

The algebraic error is faster to compute but the quantity which is minimized is not geometrically or statistically meaningful.

5.3. Geometric Error

This is an error function based on the sum of the geometric distances (that is, Euclidean or l2 distance) between X_i and $\hat{X}_i$ and between X_i' and $\hat{X_i}' = H\hat{X_i}'$ in the image. Here is an estimated set of true coordinates which are minimized over in computing the geometric error (GE). This is a non-linear minimization problem which is handled by algorithms such as the LM algorithm. The vector can be used as an initial seed for the LM algorithm. Thus, we may express GE as:

$$GE = d(X_i, \hat{X}_i) + d(X_i', \hat{X_i}'), \quad (4)$$

where $d(X_i, \hat{X}_i)$ is the Euclidean distance between the vectors X_i and $\hat{X}_i$ and likewise for the second term. Thus, the geometric error reduces the difference between measured and estimated image coordinates.

The geometric error is much more computationally intensive than the algebraic error. However, the geometric error is sometimes preferred for the following reasons: (a) the quantity being minimized in (4) has a meaning, (b) solutions based on this error are more stable, and (c) the solution is invariant under Euclidean transformation.

5.4. Watershed Error

The idea behind watershed error metric is the watershed segmentation algorithm [13] which is based on image visualization in three dimensions consisting of two spatial coordinates and gray level. The algorithm results in a finite number of image segments. In order to address the problem of over-segmentation, the concept of markers [12, 13] is used in watershed segmentation. This technique is applied to the overlapping area between I_1 and I_2. Then, the number of segments obtained in I_1 and I_2 are calculated and their total absolute difference is used to estimate the error between the two registered images.

6. EXPERIMENTAL RESULTS - SYNTHETIC

We generated an image I_1 consisting of 20 random points and a sequence $H_0, \ldots, H_{20}$ of homographies ranging from purely affine ($t = 0$) to general projective ($t = 20$). We applied these transformations to I_1 to get $I_2^0, \ldots, I_2^{20}$. We then introduced random Gaussian noise (with standard deviation 0.01) into these images to get N_1 and $N_2^0, \ldots, N_2^{20}$. Using the SVD method described above we then compute homographies $H_a^0, .., H_a^{20}$ and $H_p^0, .., H_p^{20}$ using the affine and projective models respectively. For each of these we computed the error of the homography by comparing with the true theoretical value H_t (we took ℓ_1 distance to H_t).

The results for one run of the experiment are shown in Figure 2 below. In this graph, the x-axis represents the value of t, ranging from $t = 0$ (affine) to $t = 20$ (projective). The blue line represents the value of δ, the difference between the algebraic and geometric errors as described above. The red lines represent the errors for the homographies H_a^t and H_p^t computed using the affine and projective models (the affine error is the curve which starts out smaller).

As we see, the affine model performs better than the projective model for smaller values of t, but for $t \approx \frac{2}{20} = 10\%$ away from affine to projective the projective model begins to do better. We have done this experiment with several runs of the randomly generated object, with similar results.

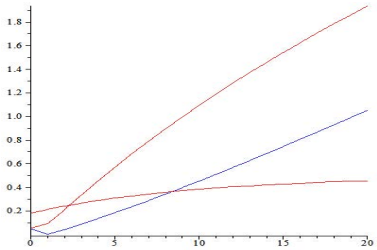


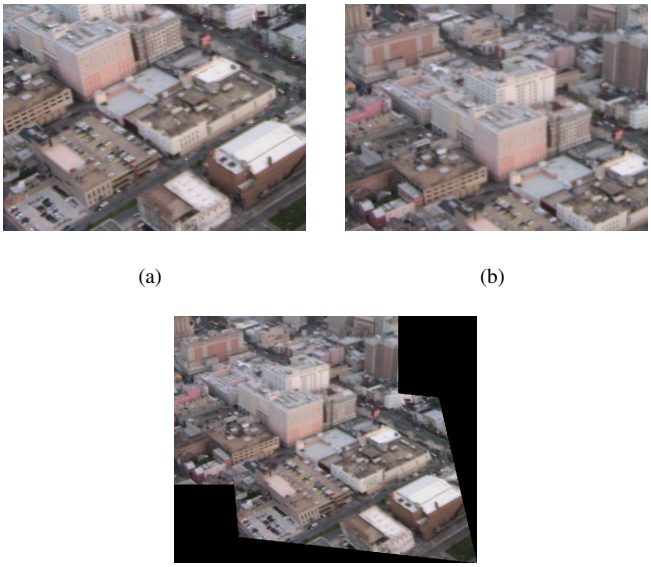
Figure 3: Results for synthetic model

7. EXPERIMENTAL RESULTS - REAL WORLD

We next ran the test on real-word images using the two different data sets mentioned previously. The camera images used for the experiment are shown below in Figure 4 and Figure 5. The images from Figure 4 represent a non-affine transformation. For this registration, it would seem best to use a general projective transformation. The images from Figure 5, on the other hand, represent an affine transformation. For both sets of images we computed the homography H assuming it represents an affine transformation, and then computed H assuming it is a general projective transformation. The homography registering the two images was computed using SIFT and RANSAC to determine corresponding points followed by the DLT method to compute the homography H (for both the affine and projective cases). The results are shown in Table 1.

7.1. Projective Transformation under Affine and Projective Homographies

From Figure 4(a) and Figure 4(b) it is clear there is a significant angle deviation between the perspectives of the two images. Thus, one would expect less error for the projective homography as opposed to the affine homography. The results below confirm that the projectively computed homography has less registration error that that of the affinely computed one.



(a) (b)

(c)

Figure 4: Real-World Images: (a) First camera image, (b) Second camera image,(c) Final Mosaic (Projective transformation)

7.2. Affine Transformation under Affine and Projective Homographies

From Figure 4 it is evident that the transformation between the two images is now essentially affine. Since the camera view in both cases is orthogonal and at at sufficiently large altitude compared to the objects in the scene, the transformation is affine. Of course, affine transformations are also projective, but by computing the homography under the assumption of affineness, there is less possibility for error in the computation of the homography (i.e., there are fewer extraneous degrees of freedom). The results are shown also in Table 1. As the results show, the error $E(A)$ for the affine homography is somewhat less than that for the error $E(P)$ for the projective homography.

8. CONCLUSIONS AND FUTURE WORK

In this paper, we investigated a method for selecting appropriate homography transformation which will reduce the error and improve the resolution of the image. We have demonstrated that such a selection improves the registration accuracy

(a) (b)

(c)

Figure 5: (a) First camera image, (b) Second Camera,(c) Final Mosaic (Affine Transformation)

Table 1: Error Results for all cases: Affine Transformation (Aff Tr) and Projective Transformation (Pro Tr) under Affine and Projective Homography

	Affine H (E(A))	Projective H (E(P))
Pro Tr (Fig 2)	20480	19370
Time (sec)	23.7	23.84
Aff Tr (Fig 3)	2906.7	2916.0
Time (sec)	3.2	3.2

and the quality of panoramic mosaic. By selecting the appropriate model for the homography subgroup, undesirable error effects can be removed. Our method is general and it can be applied to a wide range of computer vision applications for improving accuracy and avoiding computational instability. We also discussed and compared several error metrics to evaluate the registration accuracy. Our experiments suggest that geometric error captures the registration accuracy better than others.

ACKNOWLEDGEMENT

This research is partially supported by the Texas Norman Hackerman Advanced Research Program under Grant No. 0003594-00160-2009.

References

[1] P. E. Anuta. Spatial Registration of Multispectral and Multitemporal Digital Imagery using Fast Fourier Transform Techniquese. *IEEE Trans.Geosci.Electron*, l GE-8, No-4:353–368, 1970.

[2] E.D. Castro and C. Morandi. Registration of Translated and Rotated Images using Finite Fourier Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:700–703, 1987.

[3] P. Duraisamy, B. Yassine, S. Jackson, K. Namuduri, and Ye Yu. Fast automated 2-d estimation and error reduction by model selection.

[4] E. Fernandez and R. Marti. Grasp for seam drawing in mosaicing of aerial photographic maps. *J.Heuristics*, 5:181–197, 1999.

[5] H.Y.Shum and R.Szeliski. Construction of Panormic Image Mosaics with Global and Local Alignmentt. *Int. J. Comput. Vis*, 36:101–130, 2000.

[6] Yasushi Kanazawa and Kenichi Kanatani. Stabilizing image mosaicing by model selection. *Proc. SPIE*, pages 35–51, 2000.

[7] M. Kreschner. Seamline detection in color ortho image mosaicking by use of twin snakes. *ISPRS J. Photogramm. Remote Sens*, 56:53–64, 2001.

[8] C. D. Kuglin and D. C. Hines. The Phase Correlation Image Alignment Method. *IEEE Inf.Conf.Cybern.Soc., New York*, pages 163–165, 1975.

[9] Y. Lin, Q. Yu, and G. Medioni. Map Enhanced UAV Image Sequence Registration. In *CVPR*, 2007.

[10] P.Duraisamy, Yao Shen, K.Namuduri, and S.Jackson. Error analysis and performance estimation of two different mathematical methods for image registration. *Proc of SPIE*, 7799, 2010.

[11] B. S. Reddy and B. N. Chatterji. An fft-based Technique for Translational, Rotation, and Scale invarient image registration. *IEEE Trans. Image Process*, 5:1266–1271, 1996.

[12] P. V. Wie and M. Stein. A Landsat Digital Image Rectification System. *IEEE Transactions on Geoscience Electronics*, 15:130–136, 1977.

[13] A. Zisserman and R. Hartley. *Multiple View Geometry in Computer Vision*. CRC Press, 2004.