

# Automatic Fundamental Matrix Estimation (August 2008)

Pedro Davalos, *Member, IEEE*

**Abstract**—The fundamental matrix, or bifocal tensor, used in computer vision applications is a 3 by 3 matrix of rank 2, that relates corresponding points across two different viewpoints. This basic representation defines the relationship between any two images of the same scene. We will utilize this framework with a single camera in motion, and continuously automatically estimate the fundamental matrix at each step throughout the motion sequence of consecutive frames, which will assist for localization and mapping of a moving camera.

**Index Terms**—Epipolar Geometry, Fundamental Matrix, Localization, Mapping, RANSAC, SIFT

## I. INTRODUCTION

**T**HIS work focuses on the automatic estimation of the fundamental matrix, which includes the automatic selection of corresponding points across two different images of the same scene, as accurate corresponding points in the image coordinate frames are required to compute the fundamental matrix. 3-dimensional Epipolar Geometry has been extensively researched as it applies to multiple applications such as stereo vision, 3d reconstruction, structure from motion, and all the associated derivations in the computer vision field. However, our motivation is to utilize this knowledge of the relationship between two consecutive frames from a moving camera, such that it can assist solving the localization and mapping problems. Consequently, we will assume that we have a single camera in motion, with simple motion dynamics that will result in small scene changes with an approximate acquisition rate of about 1 frame per second. Furthermore, we assume that we have prior knowledge of our camera's intrinsic calibration parameters.

This document presents a proof of concept of a practical method for automatically finding the fundamental matrix between two frames from an image sequence of a camera in motion. This process can be summarized by first processing the images and finding salient feature points using SIFT descriptors, then we find potential corresponding points between the two images by listing the pairs of candidate corresponding points by selecting the closest matching SIFT point from the previous image. Finally, we select the best

corresponding points utilizing the RANSAC algorithm with the fundamental matrix correspondence equation as the model, and the correspondence error as the RANSAC cost function. Our reasoning for selecting this process is based on the stability of SIFT feature points and the robustness of the RANSAC process for finding inliers that best fit our given model.

Additional motivation for this work includes our constraint on utilizing only a single camera as a sensing device, as we are attempting to facilitate the development of a navigation system for mobile robot localization, where a single camera system allows for an extremely versatile system capable of easy deployment and integration onto any robotic platform. This versatility is primarily due to the lack of odometry data, where localization systems can be deployed as wearable sensors, handheld mobile devices, or as payloads on non-robotic platforms for auxiliary and enhanced navigation systems. However, an optional enhancement of our current system is the integration of odometry sensor data as initial parameter estimates that would facilitate the localization process by reducing computation time.

Our goal remains the design of a robust, drift free, localization system capable of operating in near real time, with only a few passive sensors that include the single camera and optional basic odometry. However, this work concentrates on the foundation for visual navigation, where we demonstrate the feasibility of autonomous fundamental matrix estimation, which will enable localization through 3-dimensional scene reconstruction.

## II. RELATED WORK

Localization and navigation are popular research topics in computer vision where significant progress and advances have been made in the recent past. Furthermore, as this document presents an innovative approach to localization by utilizing existing subcomponents. However, our main goal for this work is the automatic computation of the fundamental matrix in a robust and efficient implementation that would allow near real time computation for motion image sequences. Armangue, Pages, Salvi and Battle [1] have compared various methods for estimating the fundamental matrix between two images with different perspectives, which is essential for us to understand the options and alternatives to estimating the fundamental matrix, yet Armangue et. al. conclude that the robust techniques such as RANSAC and Least Median Squares (LMedS) provide an efficient method of removing

Manuscript received August 8, 2008. This work was supported in part by the Departments of Computer Science at Texas A&M University in conjunction with the Spacecraft Technology Center of the Texas Engineering Experiment Station (TEES).

P. Davalos is with Texas A&M University, College Station, TX 77843 USA (phone: 979-845-8768; e-mail: pedro@tamu.edu).

outliers, which is critical for an automated process. Moreover, Armangue and Salvi [2] further compare methods for computing the fundamental matrix in a more detailed fashion, where they reach the same conclusions that the Least Median Squares (LMedS) provides the best method of removing outliers, however, this technique requires the longest computation time.

Further Fundamental Matrix computation analysis has been recently conducted by Sugaya and Kanatani [3] where their goal is to achieve high accuracy estimations. However, their approach is to evaluate various numerical methods and optimization techniques that require manual selection of accurate corresponding points.

Lei, Tsui, and Hu [4] have conducted a similar study to ours, where they evaluate automatic fundamental matrix estimation utilizing the RANSAC algorithm, however, one of the key differences between Lei's work and ours is that Lei utilizes Zhang's correlation based matching method [5] to select the initial potential corresponding points between images, while we will use the SIFT technique [6],[7] for selecting salient feature points.

As we have already briefly referenced, one of the most popular techniques for finding inlying data that best fit a model is the RANSAC algorithm, which stands for "Random Sample and Consensus," initially published by Fischler and Bolles in 1981 [8]. Since then, this has become the de facto standard for eliminating outliers from noisy and imperfect measurements. Recently, there have been several modified variations of the RANSAC algorithm proposed as an attempt to make the process more efficient and more robust. One such modification is the MultiRANSAC algorithm proposed by Zuliani, Kenney and Manjunath [9], where their motivation is to have the capability of defining the data by multiple models. In other words, there are multiple error cost functions evaluated simultaneously at each iteration of the process. This enhanced capability could potentially benefit our work by using both the correspondence equation as well as the reprojection error when evaluating the quality of a given data subset.

A different modification to the RANSAC algorithm has been recently proposed by Matas and Chum, as a Randomized RANSAC [10], where this approach is capable of reducing computation times by randomizing the hypothesis evaluation during the iterations of RANSAC. This method is significant to us, as it has the potential of optimizing our process by reducing the computation time that would allow us to operate in near real time constraints for live video motion sequences.

Malis and Marchand have also conducted a study that evaluates various computer vision estimation techniques such as robust fundamental matrix computation for 3-dimensional applications [11]. Their goal is to compare the various techniques such as parameter estimation, iterative approaches, and the robust voting methods such as RANSAC that are commonly used for different applications such as correspondence determination, camera calibration, motion estimation, and fundamental matrix estimation, where their

conclusions are helpful when selecting numerical methods for specific applications of parameter estimation.

In addition to computing the fundamental matrix and finding accurate corresponding points automatically, there still remains the problem of motion estimation. Mester [12] presents a mathematical approach to modeling direction and motion estimates, where he presents extensive formulations on how to quantify motion and how to deal with measurement noise, which remains a critical component for all computer vision applications.

Further related work in the field of point correspondence has been conducted by Scharstein, Szelski, and Zabih, where they focus their research on finding the pixel correspondence for disparity computations of stereo systems or "disparity space images" [13]. Where although they are operating under different assumptions, their work is still in the field of multi-view epipolar geometry, and their findings are potentially significant for us, as it can assist us in the correspondence problem, that would allow an accurate selection of corresponding points.

Recently, Strelow and Singh [14] published their work which is closely related to ours, where they utilize SIFT feature points and the RANSAC algorithm to find corresponding points and estimate the scene's 3-dimensional structure and simultaneously determine the camera's location.

Finally, the work done by Van den Hengel, Dick, Thormahlen, Ward, and Torr relates to our work as they attempt to recover 3-dimensional structure from motion. However, they require user input for selection of rigid well defined structures, and they further assume repetitive patterns of rigid structures in the scene, thus rendering their approach incompatible with automatic correspondence determination between multiple views.

### III. FUNDAMENTAL MATRIX BACKGROUND

Epipolar Geometry refers to the relationship between multiple views or projections of the same scene when moving the center of the camera. Fig. 1 illustrates the theory behind epipolar geometry, where  $M$  is a feature point in the real world,  $m$  and  $m'$  are the corresponding locations (represented as 3-vector homogeneous coordinates) of point  $M$  in image frames  $I$  and  $I'$  respectively. Furthermore the line between the two camera centers  $C$  and  $C'$  is called the baseline, which traverses through the epipoles  $e$  and  $e'$ .

Consequently the relationship that defines the correspondence between the two images, can be defined as follows:

$$m'^T \cdot F \cdot m = 0 \quad (1)$$

Where  $F$  is the 3x3 Fundamental Matrix of rank 2 that can be estimated in the simplest form if given  $n$  corresponding points, we can reshape the  $F$  matrix as a 9-vector  $f$  and solve for  $f$  given the following relationship:

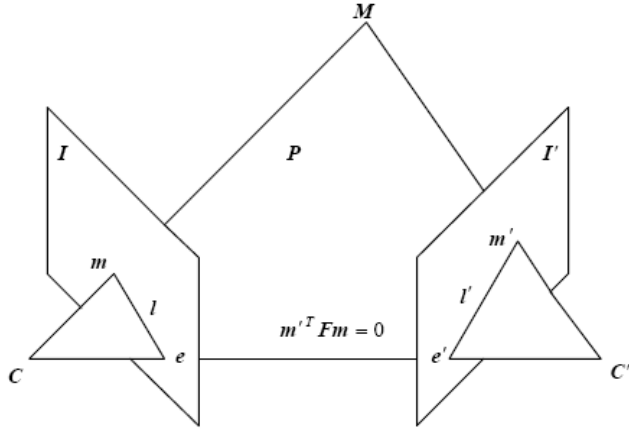


Fig. 1. Epipolar Geometry

$$A \cdot f = 0$$

with,

$$A = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} \quad (2)$$

and,

$$f = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T$$

Where  $A$  is the collection of corresponding points  $m_i = [x_i, y_i, 1]$  and  $m'_i = [x'_i, y'_i, 1]$ , with  $i=1, \dots, n$  as the index of the corresponding point. Where we can solve for  $f$  using SVD if we have at least 7 corresponding points, since the fundamental matrix has only 7 degrees of freedom.

As previously noted there are other alternative methods of computing the fundamental matrix, however, for this work we will utilize this approach with added normalization, where we convert the homogeneous image coordinates to a normalized range of values to avoid numerical instability.

#### IV. AUTOMATIC FEATURE POINT CORRESPONDENCE

Automatic feature point correspondence refers to the process of selecting points across two different images that correspond to the same point in the real world. The task of automating this process has been the topic of extensive research in the past among the computer vision field. Some of the challenges among automatic point correspondence include illumination changes, occlusion, specularity that may generate non-existing features, and the geometry perspective deformations as objects are not identical when rotated, scale deformations, among others. Consequently, automatic correspondence requires an extensive search across both images to find appropriate matchings. As previously mentioned, the fundamental matrix estimation requires an accurate set of corresponding points in order to produce a quality result. Therefore, we must solve the correspondence problem as a preprocessing step of solving for the fundamental matrix. This automatic selection of

corresponding points constitutes our noisy measurements, since we will always have to deal with imperfect selection. Yet, the simple formulation to solve for the fundamental matrix in (2) does not tolerate mismatches in the data.

In order to eliminate outliers, or mismatched points, from our data, we will utilize the RANSAC algorithm for a robust estimator, which will use the correspondence equation (1) as the model to try to fit the data. We will use an open source implementation of RANSAC published by M. Zuiani [16], however, Zuiani's implementation does not provide for a fundamental matrix estimation. Consequently, we modified the source code accordingly to incorporate the model for fitting corresponding points to the basic fundamental matrix relationship (2).

#### V. RESULTS

For this work we conducted an experiment by processing image pairs of various scenes. Where each image pair contains different perspectives of the same scene, Fig. 2 illustrates an image pair of the same scene with the camera centers at different positions.



Fig. 2. Image pair of same scene with different camera center positions.

All simulations for this work were conducted using Matlab on a workstation with a Pentium Core2 Duo processor at 2.6 GHz, and 4GB of memory.

After processing the two frames where each image has a size 120x160 pixels, we first extract SIFT features for both images. The first image (Fig 2. left) resulted in 181 sift features, and the second image (Fig. 2 right) resulted in 166 sift features. The next step is to select the image with fewer sift features, and find the most similar feature from the other image, this matching can be thresholded to a given similarity value. This process generates our list of potential candidates of corresponding points that will be processed by the RANSAC algorithm.

Fig. 3 shows the output of RANSAC, where the optimum corresponding points are shown with green dots, and the mismatched outliers are shown as red asterisks. This result illustrates the optimum consensus set, shown as green dots that optimized the fundamental matrix correspondence relationship from our given set of input data points.

After running this RANSAC process, the final consensus set of inliers, is evaluated using the fundamental matrix correspondence relationship and each pair of corresponding points is evaluated using (1), where the result should produce a zero mean distribution. Fig. 4 shows a histogram of the

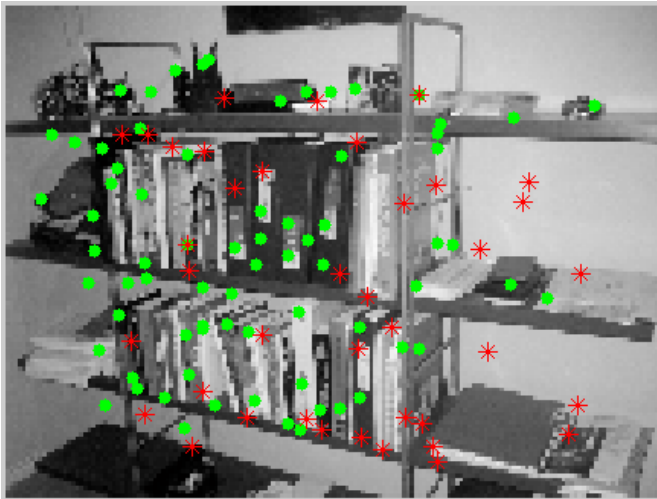


Fig. 3. RANSAC Output: Consensus Set (93 green dots) and outliers (red asterisks)

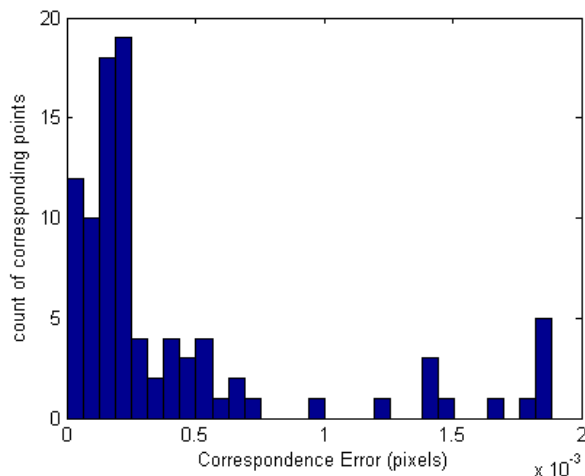


Fig. 4. RANSAC Output: Distribution of Pixel Correspondence Error using the consensus set and final Fundamental matrix selected by RANSAC. Ideally zero for all corresponding points.

distribution of pixel corresponding point error, where the values are close to zero as expected.

Finally, we verify and validate the fundamental matrix produced by RANSAC, by manually selecting accurate corresponding points across the input images from Fig. 2 and evaluating the correspondence error as specified by (1) and the pixel reprojection error. The distribution of correspondence error is shown in Fig. 5, where the negligible error indicates an accurate fundamental matrix estimate. Fig. 6 illustrates the distribution of the pixel reprojection error using the manually selected corresponding points and the RANSAC calculated fundamental matrix, which also indicates an accurate fundamental matrix estimate by the small reprojection error.

Furthermore, using the RANSAC process should include consideration of the non-deterministic algorithm, which results in a different output each time due to the random nature of RANSAC. Therefore, we will briefly analyze the performance given the same input through statistical analysis,

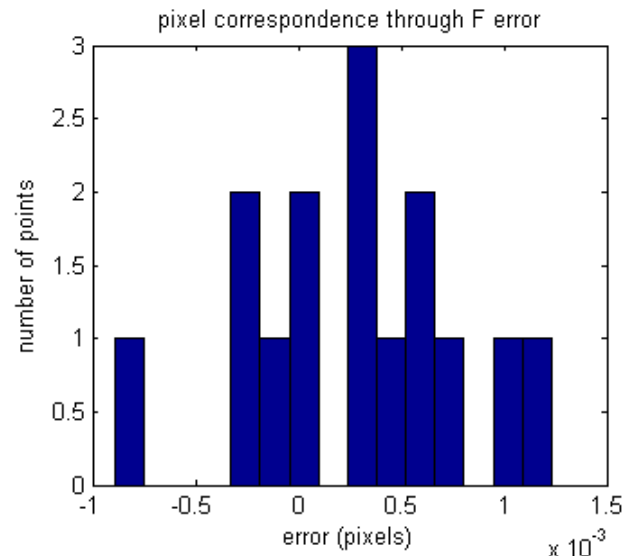


Fig. 5. Manual Verification: Distribution of Correspondence Error from RANSAC selected Fundamental Matrix and manually selected corresponding points.

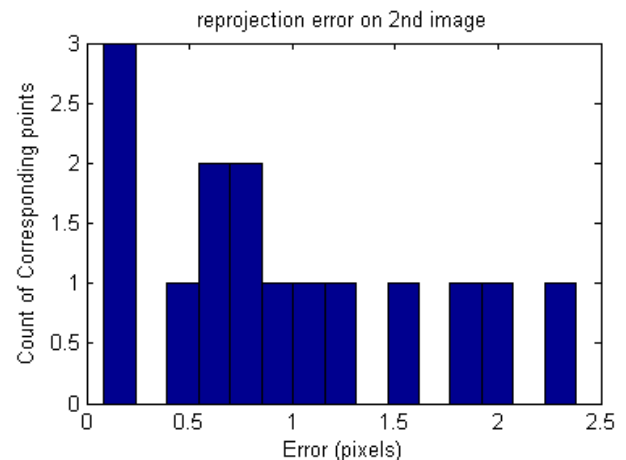


Fig. 6. Manual Verification: Distribution of Pixel Reprojection Error, from RANSAC selected Fundamental Matrix and manually selected corresponding points.

by processing the same input frames multiple times, and evaluating RANSAC's processing time and mean error. Fig. 7 shows the average processing time after 100 test cases of the same input. This figure illustrates how over 70% of the time, RANSAC was able to produce an accurate estimate in under 1 second of processing time.

Fig. 8 complements our statistical analysis by showing the distribution of mean correspondence error in pixels of each RANSAC run, across 100 test runs. This figure confirms that RANSAC is consistent in generating accurate fundamental matrix estimates.

## VI. FUTURE WORK

As we have currently demonstrated, the automatic computation of corresponding points and the continuous automatic estimation of the fundamental matrix is feasible in

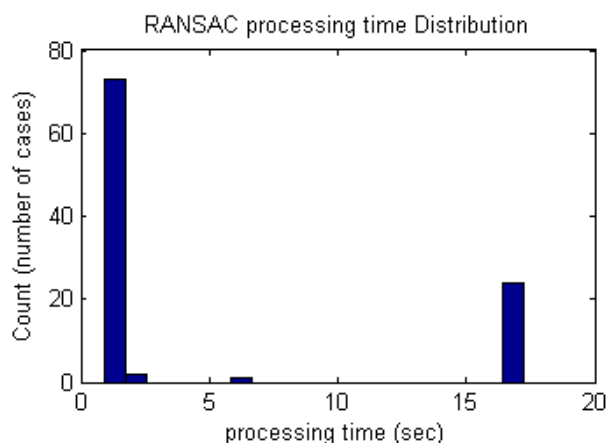


Fig. 7. RANSAC Statistical Analysis of RUNTIME.

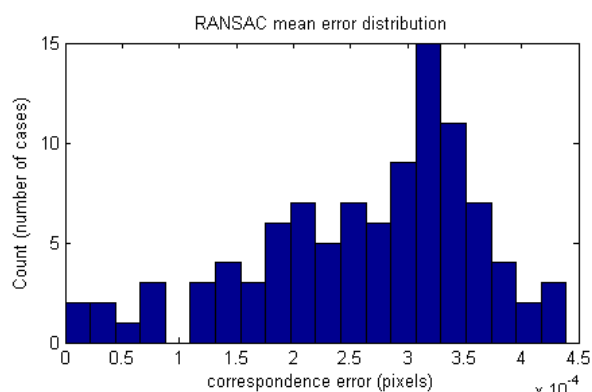


Fig. 8. RANSAC Statistical Analysis of Mean Error from correspondence error through the Fundamental Matrix.

near real time operation. However, we believe that odometry sensor data such as wheel encodings or acceleration data from an inertial measurement unit (IMU) can be incorporated to the algorithm as initial position estimates, allowing faster and more efficient computations of the correspondence and fundamental matrix that would also allow for more accurate estimations.

Further work also includes sensitivity analysis to evaluate the impact of noise in the measurements, as well as lack of distinctive patterns such as blank walls, or lack of out-of-plane features such as images of a flat wall or picture that may result in degenerate cases of the fundamental matrix estimation. Furthermore, we are yet to analyze the fundamental matrix to extract position changes.

Finally, further work is still pending to pursue the incorporation of an obstacle detection and avoidance system, as we are currently continuously estimating the 3d geometry of the observed scene, we could easily detect any feature points that are potential threats or obstacles in our given trajectory or path.

## VII. CONCLUSION

With this work, we have successfully demonstrated that we

can efficiently compute the fundamental matrix in a continuous mode of operation as desired. This work was motivated by our goal of autonomous navigation with minimal sensor requirements, as we are currently depending exclusively on a single camera system. However, as previously mentioned, we intend on pursuing further research on evaluating the integration of additional odometry data through sensor fusion for an enhanced navigation and localization system. This integration of vision plus odometry would overcome many of the issues associated with single camera systems, such as the temporary loss of vision while in motion.

It should be noted that our current work, although complete and operational, is intended as a proof of concept, with vast room for performance optimization. However, we believe, that with the appropriate effort, and with near future generation hardware, a real time implementation of our current interpretation can become practical.

Furthermore, we have also learned that SIFT features provide an extremely robust method for extracting feature points. This robustness from SIFT allows the selection of a sufficient number of correct corresponding points that allow an efficient operation of RANSAC.

## ACKNOWLEDGMENT

A special thanks to Dr. Dezhen Song for his critical role as instructor, mentor, advisor, and contributor to this study.

## REFERENCES

- [1] X. Armangué and J. Pagès and J. Salvi and J. Batlle, "Comparative Survey on Fundamental Matrix Estimation" *IX Simposium Nacional de Reconocimiento de Formas y Análisis de Imágenes*, SNRFAI 2001, pp 227-232, Castelló, Spain, May 2001.
- [2] X. Armangué, J. Salvi, "Overall View Regarding Fundamental Matrix Estimation," *Image and Vision Computing* 21(2), pp 205-220, February 2003
- [3] Y Sugaya and K Kanatani, "High accuracy computation of rank-constrained fundamental matrix by efficient search" *IEICE Transactions on Information and Systems*, Vol. E90-D, No. 2 (2007-2), pp. 579-585.
- [4] C. Lei, H. T. Tsui, and Z. Y. Hu, "On the Automatic Estimation of Fundamental Matrix" *In Proc.the 5th Asian Conference on Computer Vision*, Melbourne, Australia, Vol.2, pp.664-669,2002.
- [5] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, 75(1-2):87—120, 1995.
- [6] [15] David G. Lowe, "Object recognition from local scale-invariant features," *International Conference on Computer Vision*, Corfu, Greece (September 1999), pp. 1150-1157
- [7] [16] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.
- [8] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography" *Communications of the ACM archive* Volume 24 , Issue 6. June 1981.
- [9] M Zuliani and C S Kenney and B S Manjunath, "The multiransac algorithm and its application to detect planar homographies," *Proc. IEEE International Conference on Image Processing*, Genova, Italy, Sep. 2005.
- [10] Jir Matas and Ondrej Chum, "Randomized ransac with sequential probability ratio test," *Proc. IEEE International Conference on*

*Computer Vision (ICCV)*, volume II, pages 1727-1732, New York, USA, October 2005

- [11] E. Malis, and E. Marchand, "Experiments with robust estimation techniques in real-time robot vision," *IEEE/RSJ International Conference on Intelligent Robots Systems*, Beijing, China, 2006.
- [12] R. Mester, "On the mathematical structure of direction and motion estimation," *3rd International Symposium on Physics in Signal and Image Processing (PSIP)* Grenoble, France January 29-31, 2003
- [13] D. Scharstein, R. Szeliski, R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, 2002
- [14] D. Strelow and S. Singh, "Long-term motion estimation from images" in *Proceedings, International Symposium on Experimental Robotics*, 2006
- [15] A. van den Hengel, A. Dick, T. Thormählen, B. Ward, and P. Torr, "Building models of regular scenes from structure and motion" In *Proceedings British Machine Vision Conference*, 2006.
- [16] M. Zuliani, "Ransac Toolbox" [vision.ece.ucsb.edu/~zuliani](http://vision.ece.ucsb.edu/~zuliani)

**Pedro Davalos** obtained his bachelor's degree in computer engineering (2002) and is currently pursuing a masters degree in computer science, both at Texas A&M University, College Station, TX 77843, USA.

Mr. Davalos has worked on various projects as a Research Engineer at the Spacecraft Technology Center since 2002 (College Station, TX). Previous positions also include Senior Video Network Specialist at the Educational Broadcast Services Department of Texas A&M. Where his research interests include signal processing, machine learning, robotics, and computer vision. Mr. Davalos is also a member of the American Institute of Aeronautics and Astronautics.