A Variable Precision Hybrid Camera Calibration Method

Alison Luo and Alex Pang
Computer Science Department
University of California, Santa Cruz, CA, USA
email: {alison, pang}@cse.ucsc.edu

ABSTRACT
We propose a hybrid camera calibration based on CVP (calibration using vanishing points) and BFC (brute-force calibration) that can produce solutions of varying precision. The application context that we are looking at requires matching a 3D model of known dimensions to a single image of the object. As such, our algorithm only requires a single image of a cuboid object (representing the bounding box) as input. CVP is a closed-form calibration method and provides a quick initial estimate. However, it gives larger errors when applied to low resolution images. BFC is an iterative solution that progressively provides higher accuracy with each successive refinement. But BFC, by itself, is very time consuming because it has to search through a large space of possible transformations of the model. We therefore propose to apply CVP first to obtain a coarse estimate of the transformation and then apply BFC to refine the solution. This can be viewed as either improving the CVP accuracy or speeding up BFC. In this paper, we also provide comparisons of the accuracy of four different calibration techniques: CVP, BFC, HC (hybrid calibration), and Tsai's methods. The results show that our hybrid method is the best one in general.

KEY WORDS
Visualization, Rendering, Aeronautics

1 Introduction

The application driving this work is an integrated image versus data level comparison where 2D images are compared against 3D volumetric data. This is important in a number of situations. For example, Uselton [7] and subsequently VISOR [1], described how images of air foils with pressure sensitive paint obtained from wind tunnel experiments need to be compared against calculations of the pressure field over an equivalent geometrically modeled foil (see Figure 1). For such applications, we typically have knowledge of a 3D model, but do not necessarily have the camera parameters nor correspondence between 3D model points and 2D image points. Another application is the comparison of medical data acquired using different modalities. For example, 2D x-ray images versus a 3D volumetric CT or MRI scan. Here, fiducial marks may or may not be present.

Given such an application context, we make the following working assumptions: (1) the 3D model is a cuboid object (representing the bounding box of the object) of known size, (2) there is insignificant distortion in the image, and (3) the aspect ratio of the camera is known. With these assumptions, we examine a number of options of how extrinsic camera parameters may be derived from a single image. Because our goal is to do comparison, the accuracy of the procedure is more important than the robustness.

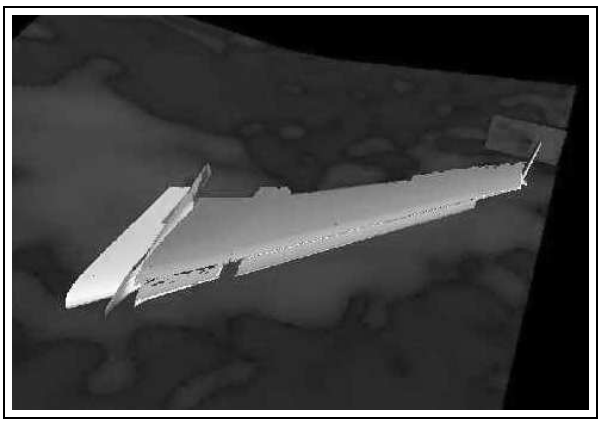

Figure 1. Image data of pressure sensitive paint superimposed on 3D wing geometry. Image courtesy of Leslie Keely, NASA.

BFC starts with a single image of a bounding box. A systematic search is then performed where different projections of a 3D cuboid model are compared against the input image. CVP uses the number and positions of vanishing points in the image to estimate camera calibration parameters. Our proposed hybrid calibration approach (HC) applies CVP first to obtain a coarse estimate of the camera transformation. It then applies BFC to iteratively improve the accuracy of the solution. As a result, HC is more accurate than CVP and much faster than BFC.

In this paper, we study four camera calibration techniques: BFC, CVP, HC, and Tsai camera calibration. We show the comparison of transformations given by these calibration techniques on our synthetic test images, and conclude that HC is the most accurate one among the four.

We introduce related work on camera calibration in Section 2, and four calibration techniques in Section 3 The results of accuracy comparisons among these techniques are given in Section 4

2 Related Work

Shen *et al.* [5] described a system for performing either image level comparison or data level comparison of wind tunnel experiments and computational fluid dynamics calculations. However, it did not support comparison between 2D images and 3D computations.

Camera calibration in this case is a bridge between 2D images and 3D volume data. The work done by Tsai [6] clearly introduced and discussed some important techniques of camera calibration which are based on the pinhole model. A 3D to 2D mapping system is accurately established by Tsai, taking into account the camera's intrinsic parameters and an uncertainty scale factor s_x. In order to solve the equations in Tsai's system, there must be a sufficient number of known 3D to 2D matched points. It is theoretically possible to compute both camera position and orientation from the 3D points and its corresponding 2D points [4]. However, one limitation is that the origin of the world space cannot be at the center of the field of view.

There are calibration techniques based on vanishing points. Wei and Song [8] showed that, if PA, PB and PC are three orthogonal edges of a cube, w_1, w_2 and w_3 are three vectors from the perspective center to the three vanishing points associated with PA, PB and PC, then PA and w_1, PB and w_2, PC and w_3 are parallel to each other respectively.

Caprile and Torre [2] demonstrated that the camera's intrinsic parameter focal length can be recovered from a single image of a cube. Furthermore, from two images of the cube, the camera's extrinsic parameters, rotation and translation from one camera to the other one, can be obtained by matching the corresponding vanishing points in the two images.

More recently, Eguillou *et al.* [3] introduced a new technique, also based on vanishing points, that extended Caprile and Torre's work. This technique produced a coarse 3D reconstruction from a single image by approximating the model with a bounding box. We refer to this method as the CVP method in our paper, but using line and corner detection instead of manual selection. Eguillou's method requires manual selection of lines in the image, which may affect the precision of recovered camera parameters. In their case, this is not a crucial problem, as their objects (bounding boxes) can be adjusted to compensate for any calibration error.

3 Camera Calibration

Every 3D model has a bounding box. Calibration based on an image of the bounding box simplifies the constraints for CVP and BFC, and makes it easier to collect 2D and 3D data points for Tsai calibration because of a regular structure of the model. The transformations of the bounding box and the transformations of the camera are complementary.

3.1 Basic Concepts

Our camera is based on a pinhole model. There are two types of camera calibration techniques: iterative solution and closed-form solution. For the iterative solution, we further assume that the camera is stationary while the model is transformed. On the other hand, for the closed-form solution, most implementations assume that the model is stationary and the image is taken from a new camera position and orientation. BFC gives an iterative solution, while CVP gives a closed-form solution. As such, BFC generates a rotation matrix from a fixed camera viewpoint for different orientations of the cuboid model. The projected model is matched against the input image to determine which model orientation has the least error (Figure 2). In contrast, CVP obtains the rotation matrix of the camera directly using the following convention: (a) vectors and points are post-multiplied by the matrix, and (b) rotation is done in the following order: x, then y, then z. That is, $[Rz][Ry][Rx]\vec{p} = [R_{xyz}]\vec{p}$. Rx, R_y and R_z are the individual rotation matrices for rotation along X, Y and Z respectively. R_{xyz} is the composite rotation matrix.

$$R_{xyz} = \begin{bmatrix} \cos z \cos y & (\cos z \sin y \sin x & (\cos z \sin y \cos x + \\ & -\sin z \cos x) & \sin z \sin x) \\ \sin z \cos y & (\sin z \sin y \sin x & (\sin z \sin y \cos x \\ & +\cos z \cos x) & -\cos z \sin x) \\ -\sin y & \cos y \sin x & \cos y \cos x \end{bmatrix} \quad (1)$$

3.2 Brute-force Calibration

Brute-force calibration is a model-driven calibration. It assumes the camera is fixed and the model is transformed in camera coordinates. There are six degrees of freedom for the model transformation. While not necessary, significant savings can be achieved if it is provided with an initial estimate of the transformation.

The search space can be reduced to $[0, 360) \times [0, 180) \times [0, 360)$, of which the first two rotation searches are in polar coordinates and the third rotation is around an axis pointing from the origin to one of the corners of the object.

For translation, we use a heuristic to limit the search space. Specifically, the ratio of the object size in the image to the size of the image gives an indication of how far away (Z translation) the object was moved, and subsequently also limits the amount of X and Y translations. Suppose the model rotation space is μ and translation space is ν, then, the BFC searching space is $\mu \times \nu$.

The precision of the solution is dependent on two parameters: r_s and t_s. For example, in the course of searching through the rotation space μ, an r_s of 10 degree will search each axis of rotation in increments of 10 degrees. Once the rotations are determined to within 10 degrees, a subsequent search of the rotation space can be initiated with

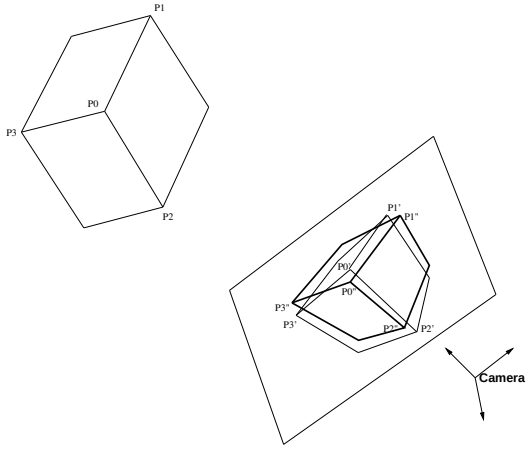

Figure 2. Image alignment between the given image (in thick lines) and the rendered image (in thin lines)

r_s of 1 degree, and so on. This multi-stage approach is more efficient than searching the entire space in 1 degree increments. BFC provides as many levels of precision as necessary with repeated refinements. To speed up computation, an initial rough estimate of the model transformation can be obtained using large refinement parameters r_s and t_s, which can be 10, 1, 0.1, 0.01.... We recommend using one-tenth of the previous values in each of the subsequent iterations.

For each possible transformation, the projected image is compared with the given image, in terms of the projected corners and the points found in the image. The error is the sum of distance difference between each pair of projected corner and image corner. Because there is ambiguity of matching the corners in the image with the corresponding corners projected from model corners, BFC requires some knowledge of the matching between projected corners and the image corners. One of the corners of the model is picked together with its three adjacent corners projected to the image plane as P'_0, P'_1, P'_2 and P'_3.

$$P'_i = RSP_i + T \qquad where\ \ i = 0, 1, .., 7 \qquad (2)$$

P_i are corners of the cuboid model. P'_is are the projections of the model corners on the image plane. R and S are rotation and scaling matrices. T is a translation vector. Among the corners detected in the image, the one (P''_0) with three edges extending from it and closest to the center of the image is picked automatically. The three corners at the other ends of the three edges emanating from P''_0 are P''_1, P''_2, P''_3. Error between the two sets of geometry is

$$error_{srt} = \sum_{i=0}^{i=3} distance(P'_i, P''_i) \qquad (3)$$

Searching through space $\mu \times \nu$, transformations resulting in the minimum error are kept for further refinement.

3.3 The Hybrid Camera Calibration Model

Camera calibration using brute-force method is constrained by the speed of refinement.

Algorithm 1 Hybrid Calibration Algorithm

Require: t_0, r_0 are initial guesses obtained from CVP
  while desired precision hasn't been met **do**
    for all 3D translations and rotations t_i and r_i **do**
      $t_i \leftarrow t_0 - t_{s-1}$
      while $t_i \leq (t_0 + t_{s-1})$ **do**
        $r_i \leftarrow r_0 - r_{s-1}$
        while $r_i \leq (r_0 + r_{s-1})$ **do**
          $error_{srt} = \sum_{i=1}^{i=m} distance(P'_i, P''_i)$
          $r_i \leftarrow r_i + r_s$
        end while
        $t_i \leftarrow t_i + t_s$
      end while
    end for
    $r_0 \leftarrow r_{min}$; $r_{s-1} \leftarrow r_s$; $r_s \leftarrow 0.1 r_s$;
    $t_0 \leftarrow t_{min}$; $t_{s-1} \leftarrow t_s$; $t_s \leftarrow 0.1 t_s$;
  end while
  Output: t_0 and r_0

CVP has higher accuracy if the image resolution is high since its accuracy is directly dependent on the accuracy of the extracted vanishing points. Therefore, its accuracy suffers if the image resolution is low or the vanishing points are not obtained accurately. However, it produces fairly good results up to a certain degree of precision and is therefore a good candidate for producing the initial estimate as input for the BFC method. As the BFC generally converges to the correct transformation, it improves the transformation from CVP to whatever precision is desired, as described in the Hybrid Calibration algorithm. t_s and r_s are the current refinement increments for translation and rotation respectively. t_{s-1} and r_{s-1} are the refinement increments of the previous stage, and are 10 times larger than the current increments. P''_i and P'_i are the corners in the given image and the corners projected from the 3D model respectively. t_{min} and r_{min} are the transformations that produce the minimum errors in the corners alignment. m is the number of corner pairs used in the alignment comparison.

4 Results

We compared four camera calibration techniques: CVP, BFC, Tsai calibration and HC using several computer generated images as test images. This provides a controlled setting for testing the accuracy of the calibration methods independent of the errors from extracting corner and vanishing points in real images where noise may be a major factor. Furthermore, for the purposes of our application of integrated data and image level comparison, the 3D data sets along with any intermediate measurements need to be rendered before they can be compared with 2D images. It therefore makes sense to perform the comparison using computer generated images. A subset of our results using

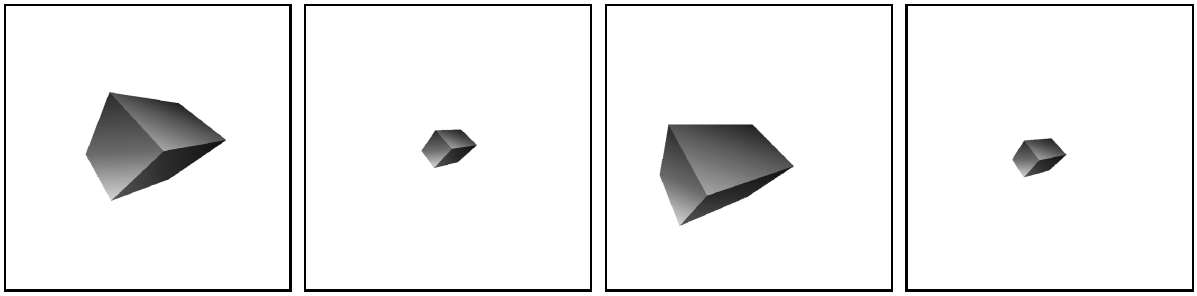

Image 1	Image 2	Image 3	Image 4
R(+56.66,-41.20,-10.30)	R(+56.66,-41.20,-10.30)	R(56.66,-41.20,-10.30)	R(+56.66,-41.20,-10.30)
T(+0.00,+0.00,+2.51)	T(+0.00,+0.00,+6.31)	T(-0.50,+0.50,+2.51)	T(-0.50,+0.50,-6.31)

Figure 3. Rotation and translation vectors for near and far objects without (left pair) and with (right pair) XY translations.

four images are presented, which are shown in Figure 3 together with the extrinsic camera parameters used to generate them. Image 2 and 4 are similar to image 1 and 3 respectively, but with the object farther away. Due to the errors introduced by the corner and vanishing point detection, calibration results based on image 2 or 4 have less accuracy than their counterparts with higher resolution objects. Image 3 and 4 are off centered versions of image 1 and 2 respectively. The former pair is used in Tsai's calibration since it cannot handle objects centered along the line of projection.

The accuracy check we performed proceeds by generating a perspective projection of the model first with openGL. The transformations applied to this object are recorded and used to measure how accurately the different methods can estimate it. We refer to this as the target camera parameters. Each algorithm is run twice against two conditions. The first is with corner and/or vanishing points extracted using image processing techniques from the 2D images. Here, we apply an edge detector on the input image. Hough transformation is then used to represent the lines found in the image. In order to locate the corners, each Hough line is traced. The two end points of each line are candidates for corner points. The corner candidates are clustered using a user-defined threshold. In the second accuracy check, the 3D corner points are projected using the exact transformations. These projected corner points (and correspondingly, the theoretical locations of the vanishing points) are then used as input for the calibration routines. From these two conditions, we derive the extrinsic camera parameters. We then use the extracted camera parameters to reproject the 3D cuboid model, and compare the projections against those obtained by using the target camera parameters.

In BFC the rotation search in cartesian space is simply $[0, 360) \times [0, 360) \times [0, 360)$, though it has potential solutions that are not unique, which means this search space is not optimal and it is very costly to go through for BFC. For these reasons, in our experiments we used the rotation search space $[-90, 90] \times [-90, 90] \times [-90, 90]$. The translation search space choosen is $[0, 1] \times [0, 1] \times (2, 7]$. These settings are based on initial guesses. If they are not set appropriately, the calibration results will not be effective. This strenghens the need for CVP as an initial guess method.

Tables 1 - 4 summarize the test results for each algorithm against (a) detected points from processing the images, and (b) "perfect" points provided to the algorithms by projecting the corners of the cuboid using the extrinsic camera parameters in Figure 3. We use the L_2 norm to express the errors in rotation and translation.

5 Conclusion and Future Work

In our experiments, we have found that both CVP and Tsai calibrations give very good results, often within 1% error. This can be further improved using our proposed hybrid approach where BFC is used to iteratively refine the solution to some arbitrary precision.

Both BFC and CVP are as accurate as the input corners and vanishing points. When the image size or resolution of the target object is small or low, the accuracy of the corner or vanishing point detection suffers, and thereby leads to inaccurate camera calibration. Therefore, while CVP is theoretically accurate, in practice, it is highly sensitive to the accuracy of the input vanishing points. We also note that inaccuracies may arise from noise in the corner detection algorithms and also from the calibration algorithm itself. HC focuses on improving the latter problem.

An easy improvement to the hybrid calibration is to have an adaptive termination criteria, as opposed to some preset precision level. This could be achieved, for example, by checking the amount of improvement in the solution.

The results presented in this paper show that given a 3D scientific data set with information about its bounding box, it is feasible to apply the proposed hybrid calibration method to register 3D data sets to 2D images. This allows us to move on to our next task of performing integrated image and data level comparison.

Corners	Images	f	R	T	Error R	Error T
Detected	Image 1	+1.23	R(+54.73,-40.47,-10.04)	T(-0.02,-0.03,+2.53)	R +3.04	T +0.04
	Image 2	+1.06	R(+45.59,-35.85,-10.43)	T(-0.13,-0.17,+5.64)	R +12.31	T +0.70
Perfect	Image 1	+1.30	R(+56.66,-41.24,-10.30)	T(+0.00,+0.00,+2.51)	R +0.04	T +0.00
	Image 2	+1.30	R(+56.57,-41.37,-10.28)	T(+0.01,+0.00,+6.21)	R +0.19	T +0.10

Table 1. Calibration using vanishing points. Note that even with perfect vanishing points, CVP does not produce an exact answer. Furthermore, it also seems to be sensitive to the size of the object in the image.

Corners	Images	f	R	T	Error R	Error T
Detected	Image 1	+1.23	R_1(+60.00,-40.00,-10.00)	T_1(+0.00,+0.00,+2.28)		
			R_2(+56.00,-41.00,-10.00)	T_2(+0.00,+0.00,+2.48)		
			R_3(+56.70,-41.40,-10.30)	T_3(+0.00,+0.00,+2.52)		
			R_4(+56.64,-41.37,-10.35)	T_4(+0.00,+0.00,+2.52)	R +0.18	T +0.01
	Image 2	+1.06	R_1(+60.00,-40.00,-10.00)	T_1(+0.00,+0.00,+5.15)		
			R_2(+57.00,-41.00,-10.00)	T_2(+0.00,+0.00,+5.26)		
			R_3(+56.00,-42.00,-10.40)	T_3(-0.01,+0.02,+5.17)		
			R_4(+56.10,-42.10,-10.36)	T_4(-0.01,+0.02,+5.17)	R +1.06	T +1.14
Perfect	Image 1	+1.30	R_1(+60.00,-40.00,-10.00)	T_1(+0.00,+0.00,+2.30)		
			R_2(+56.00,-41.00,-10.00)	T_2(+0.00,+0.00,+2.50)		
			R_3(+56.70,-41.20,-10.30)	T_3(+0.00,+0.00,+2.51)		
			R_4(+56.66,-41.20,-10.30)	T_4(+0.00,+0.00,+2.51)	R +0.00	T +0.00
	Image 2	+1.30	R_1(+60.00,-40.00,-10.00)	T_1(+0.00,+0.00,+6.30)		
			R_2(+56.00,-41.00,-10.00)	T_2(+0.00,+0.00,+6.30)		
			R_3(+56.70,-41.20,-10.30)	T_3(+0.00,+0.00,+6.30)		
			R_4(+56.66,-41.20,-10.30)	T_4(+0.00,+0.00,+6.30)	R +0.00	T +0.01

Table 2. Brute-force calibration. R_i and T_i are the rotation and translation refinement at the ith iteration. BFC produces successively better solutions after each refinement. Using perfect corners, there is virtually no further improvement to the solution after three iterations in both images 1 and 2.

Corners	Images	f	R	T	Error R	Error T
Detected	Image 3	+1.37	R(+58.60,-40.40,-12.51)	T (-0.51,+0.49,+2.67)	R +3.05	T +0.73
	Image 4	+1.30	R(+50.96,-44.39,-6.77)	T(-0.54,+0.56,+3.82)	R +7.42	T +2.61
Perfect	Image 3	+1.30	R(+56.63,-41.18,-10.29)	T (-0.50,+0.50,+2.51)	R +0.04	T +0.00
	Image 4	+1.30	R(+56.63,-41.18,-10.29)	T(-0.50,+0.50,+6.30)	R +0.04	T +0.01

Table 3. Tsai calibration. Note that even with perfect corner points, it does not produce an exact answer. Furthermore, it also seems to be quite sensitive to the accuracy of the detected corners.

Corners	Images	f	(CVP) R	(BFC) R	Error R
Detected	Image 1	+1.23	R(+54.73,-40.47,-10.04)	R_2(+57.00,-41.00,-11.00)	
				R_3(+56.20,-41.30,-10.10)	
				R_4(+56.19,-41.29,-10.03)	R +0.55
	Image 2	+1.06	R(+45.59,-35.85,-10.43)	R_2(+57.00,-41.00,-10.00)	
				R_3(+56.00,-42.00,-10.40)	
				R_4(+56.07,-42.09,-10.39)	R +1.07
Corners	**Images**	**f**	**(CVP) T**	**(BFC) T**	**Error T**
Detected	Image 1	+1.23	T(-0.02,-0.03,+2.53)	T_2(+0.00,+0.00,+2.42)	
				T_3(+0.00,+0.01,+2.45)	
				T_4(-0.00,+0.01,+2.45)	T +0.06
	Image 2	+1.06	T(-0.13,-0.17,+5.64)	T_2(+0.00,+0.00,+5.20)	
				T_3(-0.01,+0.02,+5.16)	
				T_4(-0.01,+0.02,+5.17)	T +1.14
Corners	**Images**	**f**	**(CVP) R**	**(BFC) R**	**Error R**
Perfect	Image 1	+1.30	R(+56.66,-41.24,-10.30)	R_3(+56.70,-41.20,-10.30)	
				R_4(+56.66,-41.20,-10.30)	R +0.00
	Image 2	+1.30	R(+56.57,-41.37,-10.28)	R_3(+56.70,-41.20,-10.30)	
				R_4(+56.66,-41.20,-10.30)	R +0.00
Corners	**Images**	**f**	**(CVP) T**	**(BFC) T**	**Error T**
Perfect	Image 1	+1.30	T(+0.00,+0.00,+2.51)	T_3(+0.00,+0.00,+2.51)	
				T_4(+0.00,+0.00,+2.51)	T +0.00
	Image 2	+1.30	T(+0.00,+0.00,+6.21)	T_3(+0.00,+0.00,+6.30)	
				T_4(+0.00,+0.00,+6.30)	T +0.01

Table 4. Hybrid calibration. With perfect corners, CVP provides very good initial estimates, so that exact answers are obtained after only two iterations. Initial estimates are the same as colume four and five in table 1.

6 Acknowledgements

We thank Dr. Peyman Milanfar for original discussions. We wish to thank Dr. Roger Tsai for confirming the validity of the pinhole model with his calibration technique. We would particularly like to thank Dr. Craig Wittenbrink and Dr. Mark Livingston, from Hewlett-Packard Laboratories, for their help in general, and especially for Mark's help with a more robust solution of the Tsai calibration. We would also like to thank Dr. R. G. Willson for his implementation of the Tsai calibration. Thanks to Leslie Keely from NASA for providing the image of the wing data set. This project is supported in part by LLNL Agreement No. B347879 under DOE Contract No. W-7405-ENG-48, NSF ACI-9908881, and a NASA IS grant.

References

[1] "www.nas.nasa.gov/software/msv/visor.html." VISOR: Visual Integration of Simulated and Observed Results, 1999. NASA website.

[2] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *International Journal of Computer vision*, vol. 4, pp. 127–40, 1990.

[3] E. Guillou, D. Meneveaux, E. Maisel, and K. Bouatouch, "Using vanishing points for camera calibration and coarse 3D reconstruction from a single image," *Visual Computer*, vol. 16, no. 7, pp. 396–410, 2000.

[4] R. Horaud, B. Conio, and O. Leboulleux, "An analytic solution for the perspective 4-point problem," in *Proceedings CVPR '89 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 500–507, IEEE Computer Society Press, June 1989.

[5] Q. Shen, A. Pang, and S. Uselton, "Data level comparison of wind tunnel and computational fluid dynamics data," in *Proceedings of Visualization 98*, pp. 415–418, 557, 1998.

[6] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lens," *IEEE Journal of Robotics and Autoation*, vol. RA-3, pp. 323–44, Aug. 1987.

[7] S. Uselton, "ExVis and wind tunnel experiment data," *IEEE Computer Graphics and Applications*, vol. 18, pp. 75–77, July-August 1998.

[8] G.-Q. Wei, Z. He, and S. D. Ma, "Camera calibration by vanishing point and cross ratio," in *International Conference on Acoustics, Speech and Signal Processing*, vol. 3.4, pp. 1630–3, 1989.