Motion Based Image Registration with Applications Toward Multimodal and Coaxial Camera Rigs

Paper ID ****< replace **** here and in header with paperID>

Abstract

Today's image registration techniques align pairs of images using image features or pixel intensities common between image pairs to find correspondences that are in turn used to find the geometric transformation that allows the sensed image to be warped into the reference image. However, some computer vision applications that require image, don't produce the desired results using correspondences derived from image features or pixel intensities. Two examples are the multimodal camera rig and the center region of a coaxial camera rig. In this paper we present a novel automatic image registration technique using variational methods for aligning image sequences by aligning the optical flow fields thus eliminating the need to find common image features or intensities between the images being registered. Our method applies to application where there is inherent motion between the camera and the scene and where the scene has enough visual texture to produce optical flow. We apply the technique a traditional binocular stereo rig consisting of an RGB/IR camera pair and to a coaxial camera rig. We present results for synthetic flow fields and for real images sequences.

1. Introduction

Image registration is a family of techniques used to find the transformation between two images that results in the alignment of the two images [8]. Image registration is used in many image processing and computer vision applications [9, 18]. It is a key step in 3D reconstruction from multiple view applications and in many image processing applications.

Existing image registration techniques are based on matching pixel intensity values or features which are derived from pixel intensity values [8, 18, 25]. The resulting correspondences are used to find the transformation that allows one image, the sensed image, to be warped into the second image, the reference image, such that corresponding points in the scene are the same pixel in the reference and transformed image pair.

There are computer vision applications, however, where traditional image registration techniques do not produce the desired results. Two notably cases are multimodal camera rigs where the images produced from different sensor types are not similar enough to be registered using pixel intensities or features [25] and the center region of a coaxial camera rig [16] where the disparity is too small to produce good triangulation. There are also multi-camera applications where it is desirable to augment the use of pixel intensities and/or image features to improve the finding of intra-camera correspondences.

In this paper we introduce a novel automated method for registering images using the optical flow fields from two cameras and we apply the technique to images acquired by both a multimodal stereo rig where one camera contains an RGB sensor and the other camera contains an IR sensor and to a coaxial camera rig. In applications where there is sufficient motion between the camera rig and the scene (scanning security camera, camera mounted on a vehicle, endoscope, etc.) and where the scene exhibits enough texture to produce optical flow, our method finds the transformation between multi-view image sequences without using intra-camera pixel intensities or features and produces disparity maps with accuracies similar to, and in certain cases, substantially better than, techniques that align images based on image features or pixel intensities.

2. Related Work

2.1. Multimodal Binocular Stereo

Registering images from stereo rigs consisting of cameras with multimodal sensors has been an active research area for the last decade and a half. Initially inspired by the work done to match medical images to models [22] it has more recently been motivated by the need for surveillance systems that use a combination of visible light and infrared cameras to detect targets. As noted by Yaman and Kalkan [23], traditional registration techniques used in stereo vision are not applicable to multimodal camera rigs because the pixel intensities can be substantially different in a visible light image vs. an IR image. Solutions to the multi-modal problem fall into two broad categories. The first uses Mutual Information (MI).

MI was original proposed by Viola and Wells [22] to match medical images to models. To our knowledge, Egnal [4] was the first to use MI as a similarity measure to match multi-modal stereo images. Since then, numerous improvements have been made including adaptive windowing [5], incorporating prior probabilities [6], regions of interest [11-13], and extending MI using gradient information [3].

More recently, local self similarity (LSS), originally used in template matching, was proposed for use in a multi-modal camera rig [20]. Most recently Yaman and Kalkan [23] used MI to generate dense disparity maps from multi-modal camera rigs.

The method we present, avoids using visual similarity measures between the images from the two different sensor types by computing the optical flow fields from the two sensors and then aligning the flow fields. This permits images with no common features to be aligned as long as there is motion between the camera and the scene and the scene has enough texture to produce optical flow.

Verri and Poggio [21] have shown that in many cases optical flow is not equivalent to the motion field. While optical flow algorithms have improved substantially since the Verri and Poggio paper (see [17] and [19] for summaries of the progression of optical flow development); optical flow errors caused by the aperture problem, non-Lambertian surfaces, and non-uniform changing illumination, still exist.

For image registration, however, the optical flow fields do not need to be equivalent to the motion fields. For example, errors caused by the aperture problem where only the motion tangential to edges is detected or errors caused by moving shadows, will be perceived by the two sensors identically and alignment is unaffected.

2.2. Depth from Zooming - the Coaxial Camera Rig

Depth from images taken at different focal lengths along a common optical axis was first proposed by Ma and Olsen [16]. Lavest et. al. [15, 14] provide a proof for inferring 3D data from images taken coaxially at multiple focal lengths and models a revolving object using images taken at multiple focal lengths. Asada et al. [1] and Baba et al. [2] present a method for doing 3D reconstruction using blur from zoom. Gao et al. [7] present a distance measurement system for mobile robots using zooming. Most recently, Zhang and Qi [24] describe a method for 3D reconstruction from multi-focal length coaxial images uses a snake-search algorithm.

The primary reason researchers have focused on using a single camera at different focal lengths to do 3D reconstruction has been cost. However, there are several other advantages. Ma and Olsen alluded to the fact that a depth from zoom camera exhibits substantially smaller occlusions than an equivalent binocular stereo camera rig. Additionally, there are applications where a stereo baseline is prohibitive (endoscope or bore scope) and were the known correspondence point on the optical axis is an advantage to image registration. Finally, where image registration is the objective of the application (e.g. multimodal camera rigs), a coaxial camera produces substantially smaller disparity errors than a binocular stereo rig.

The coaxial camera rig [10] is equivalent to simultaneous depth from zooming, but instead of changing the focal length of a single fixed camera, two cameras are arranged such that the cameras form images along the same optical axis. This is done by splitting the optical path with a beam splitter and aligning the two cameras such that their optical centers image the same point in the 3D scene. The coaxial camera rig combined with motion based image registration overcomes the two main problems of depth from zooming. First, simultaneous images taken at two different focal lengths overcomes the stationary scene constraint of depth from zooming. Second, using the flow field to align image pairs overcomes the unrecoverable point problem in the center region described by Ma and Olsen.

3. Variational Models

3.1. Binocular Stereo

Referring to Figure 1, let $\bar{x}_l \coloneqq (x_l, y_l)^T$, $\bar{x}_r \coloneqq (x_r, y_r)^T$ represent points in the image domain of the left and right cameras. Let $\bar{h}(\bar{x}) \coloneqq$ the disparity between $\bar{x}_l$ and $\bar{x}_r$ such that $\bar{x}_l$ and $\bar{x}_r + \bar{h}(\overline{\bar{x}_r})$ represent the same point $\bar{X}(\bar{x}_l) \coloneqq (X, Y)$ in the scene. Let $f \coloneqq$ the focal lengths of the cameras and $Z_0(\bar{x}_l), Z_1(\bar{x}_l) \coloneqq$ the distance between the optical center of the left camera and a point in the scene correspond to $\bar{x}_l$ at time = 0 and 1, the distance being measured along the optical axis. $\Delta Z(\bar{x}_l)$ is then the difference in Z between the time = 1 and time = 0. $\bar{X} \coloneqq$ the distance from the optical axis to a point on the scene and $\overline{\Delta X} \coloneqq$ the change in the distance from the optical axis between time = 0 and time = 1. $b \coloneqq$ the stereo baseline. $\bar{w}_l, \bar{w}_r \coloneqq$ the ideal flow fields in the left and right cameras.

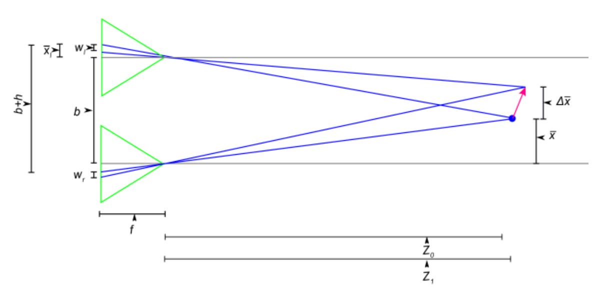

Figure 1

The relationship between an ideal flow field in the two cameras is:

$$p(\bar{x}_l)w_l(\bar{x}_l) = w_r(\bar{x}_l + h(\bar{x}_l)) \tag{1}$$

where:

$$p(\bar{x}_l) = 1 + \frac{\Delta Z(\bar{x}_l)h(\bar{x}_l)}{f\Delta\bar{X} - \Delta Z(\bar{x}_l)\bar{x}_l} \tag{2}$$

$p(\bar{x}_l)$ has a direct physical interpretation. From equation (2) it can be seen that $p(\bar{x}_l) = 1$ if $\Delta Z(\bar{x}_f) = 0$. Referring to figure 1, one can see that a change in Z introduces a slight parallax (ρ) in the finishing points of the optical flow detected by the two cameras. $p(\bar{x}_l)$ compensates for the parallax and can be solved for directly from the coaxial camera geometrically.

The first term in our stereo variational model is an optical flow matching term:

$$E_{match} = \int_a^b \frac{1}{2}[p(\bar{x}_l)w_l(\bar{x}_l) - w_r(\bar{x}_l + h(\bar{x}_l))]^2\, dx \tag{3}$$

The second term is a smoothness term:

$$E_{smooth} = \frac{1}{2}\int_a^b \|\nabla Z(\bar{x}_l)\|^2\, dx \tag{4}$$

The total energy that we want to minimize is:

$$E_{total} = \gamma E_{match} + \alpha E_{smooth} \tag{5}$$

where γ and α are tuning constants.

3.2. Coaxial Camera

Referring to Figure 2, let $\bar{x}_f \coloneqq (x_f, y_f)^T$, $\bar{x}_b \coloneqq (x_b, y_b)^T$ represent points in the image domain of the front and back cameras. Let $\bar{h}(\bar{x}) \coloneqq$ the disparity between $\bar{x}_f$ and $\bar{x}_b$ such that $\bar{x}_f$ and $\bar{x}_b - \bar{h}(\bar{x}_f)$ represent the same point $\bar{X}(\bar{x}_f) \coloneqq (X, Y)$ in the scene. Let $f_f, f_b \coloneqq$ the focal lengths for the front camera and back cameras and $Z(\bar{x}_f) \coloneqq$ the distance between the optical center of the front camera and a point in the scene correspond to $\bar{x}_f$, the distance being measured along the optical axis. $b \coloneqq$ the distance between the optical center of the two cameras. $\bar{w}_f, \bar{w}_b \coloneqq$ the ideal flow fields in the front and back cameras.

The relationship between an ideal flow field in the two cameras is:

$$m(\bar{x}_f)w_f(\bar{x}_f) = c(\bar{x}_f)w_b\left(\bar{x}_f m(\bar{x}_f)\right) \tag{6}$$

where:

$$m(\bar{x}_f) = \left(\frac{f_b}{f_f}\right)\left(\frac{Z(\bar{x}_f)}{(Z(\bar{x}_f)+b)}\right) \tag{7}$$

and

$$c(\bar{x}_f) = \left(\frac{w_f(\bar{x}_f)}{\left(\frac{Z_0(\bar{x}_f)+b}{Z_1(\bar{x}_f)+b}\right)\left(\frac{Z_1(\bar{x}_f)}{Z_0(\bar{x}_f)}\right)(w_f(\bar{x}_f)+\bar{x}_f) - \bar{x}_f}\right) \tag{8}$$

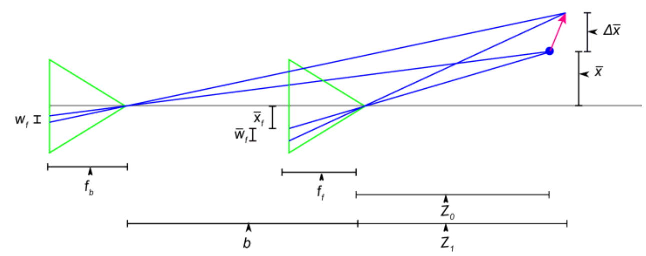

Figure 2

Where $Z_0(\bar{x}_f)$ and $Z_1(\bar{x}_f)$ are Z at time = 0 and time = 1.

Like $p(\bar{x}_l)$ in the binocular stereo example, $c(\bar{x}_f)$ has a direct physical interpretation. From equation 5 it can be seen that $c(\bar{x}_f) = 1$ if $Z_0(\bar{x}_f) = Z_1(\bar{x}_f)$ or when $\Delta Z(\bar{x}_f) = 0$. Referring to figure 3 one can see that a change in Z introduces a slight parallax (ρ) in the finishing points of the optical flow detected by the two cameras. $c(\bar{x}_f)$ compensates for the parallax and can also be solved for directly from the coaxial camera geometrically.

The first term in our coaxial camera variational model is an optical flow matching term:

$$E_{match} = \int_a^b \frac{1}{2}\left[m(\bar{x}_f)w_f(\bar{x}_f) - c(\bar{x}_f)w_b\left(\bar{x}_f m(\bar{x}_f)\right)\right]^2 dx \tag{9}$$

The second term is a smoothness term:

$$E_{smooth} = \frac{1}{2}\int_a^b \left\|\nabla Z(\bar{x}_f)\right\|^2 dx \tag{10}$$

The total energy that we want to minimize is:

$$E_{total} = \gamma E_{match} + \alpha E_{smooth} \tag{11}$$

where γ and α are tuning constants.

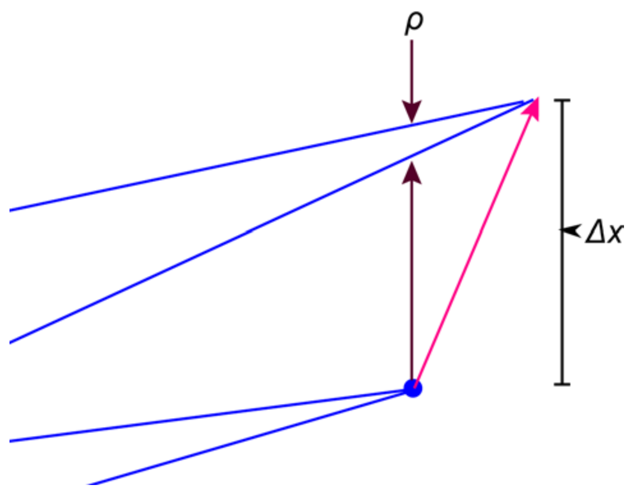

Figure 3

4. Numerical Solution

4.1. Euler-Legrange

The Euler-Lagrange equations for (3), (4), (9), and (10) are taken with respect to z. The solutions are straightforward and are not be presented here.

We reduce the problem to a 1D optimization problem by observing that the solutions for the stereo camera rig lie on horizontal epipolar lines and for the coaxial camera rig the solutions lies on radial epipolar lines. For the coaxial camera rig, we resample the front and back images onto polar coordinates. In both cases the Euler-Legrange equations (one for the x direction and one for the y direction) are solved using the gradient decent method.

4.2. Initialization

For the first iteration, we need an initial estimate for Z_0. While it's possible to solve for Z_0 and Z_1 simultaneously the best results are obtained if the first image pair only contains ΔX or only ΔZ translation. In this case either $Z_0 = Z_1$ or $\bar{X}_0 = \bar{X}_1$ and the initial estimate is significantly simplified.

For a stereo camera rig and a scene without an occlusion in the center region there exists a pixel pair on each epipolar line where $\bar{x}_l = -\bar{x}_r$. These pixel pairs represent stereo correspondences and can be used to derive an initial estimate for Z using the standard stereo equation. The optical flow can then be used to estimate the depth map along each epipolar line.

For a coaxial camera, an initial Z estimate can be made by observing that when an optical flow vector ends on the optical axis we have a special case where:

$$w_f\left(\bar{x}_f\right) = \bar{x}_f \tag{12}$$

The initial Z value can then be found using equation:

$$Z\left(\bar{x}_f = (0,0)^T\right) = \frac{b}{\frac{w_f(0) f_b}{w_b(0) f_f} - 1} \tag{13}$$

Like in the stereo configuration, we use the optical flow to make a Z estimate along each epipolar line.

4.3. Stopping Criteria

We used one of two stopping criteria depending on the quality of the flow fields and the value chosen for α. When the flow fields closely represent the motion fields and α is small (minimal Z smoothing), we use the error in the first term of the energy equation, which represents the mismatch in registration of the two flow fields, and stop when this number becomes suitably small.

Where the flow fields are noisy and not as good a representation of the motion field we need to increase α to get good results. With more substantial smoothing, the smoothing term, can pull the Z estimate away from the correct value if γ is large and/or if many iterations are performed. In this case we stopped the iterations when the smoothing term was approximately equal to, but of opposite sign to the flow matching term. This later approach often results in a larger error between the flow fields but our experiments suggest that it results in more accurate alignment because we stop iterating before the smoothness term pulls the estimate too far from the ideal solution.

5. Experiments

5.1. Synthetic optical flow field

For the synthetic optical flow fields we defined the geometry of a 3D scene and project the 3D motion of that scene onto a virtual image plane via an ideal pinhole camera model. This results in a simulated optical flow field that is exactly equal to the motion field. The simulated flow field experiments provide an estimate of the upper boundary of accuracy for our methodology and expose limitations on the 3D velocity with respect to the camera geometry.

5.1.1 Stereo Camera Rig

To determine the accuracy of the resulting image registration we reconstruct the depth map along a horizontal epipolar line using the results of registration and compare the reconstructed depth map with the original scene geometry computing both the RMS disparity error and the resulting RMS depth error.

For our synthetic flow images we created a scene geometry that ranges from 10m to 20m from the camera center. f = 4.0mm, the cameras have .006 mm square pixels, velocity in the XY plane was varied from 0.5 m/s to 3.5 m/s and velocity along the Z-axis ranged from 2.5 m/s toward the camera to 2.5 m/s away from the camera. The camera frame rate was set to 30fps. We set $\gamma = 1 \cdot 10^{9}$ and $\alpha = 1 \cdot 10^{-1}$.

Figures 4 and 5 show the results for a smooth scene without any occlusions. The worst case RMS depth error is < 0.25% and worst case RMS disparity errors < 0.01 pixels. The accuracy is slightly reduced as delta Z increases and delta X decreases. We believe that this slight reduction in accuracy is due to the cancellation that occurs in the flow fields between X and Z translations in some areas of the image.

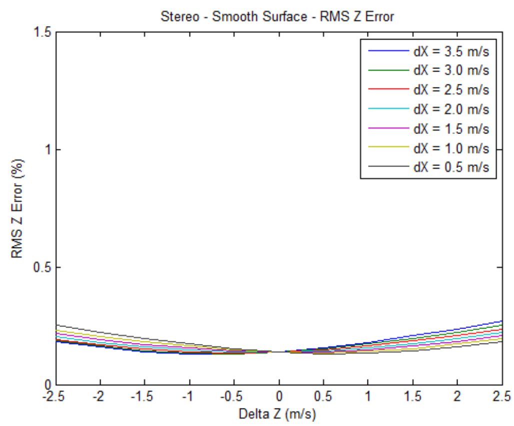

Figure 4

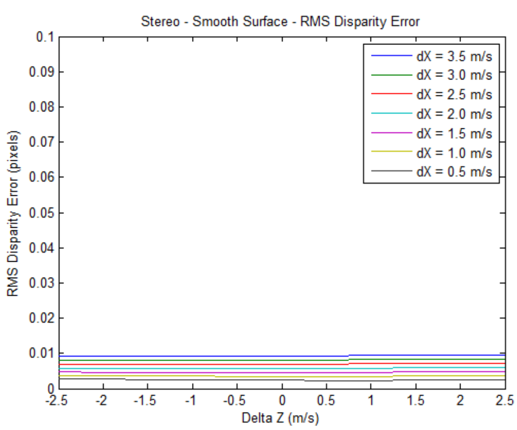

Figure 5

Figures 6 and 7 show the results for a scene with a large occlusion caused by a large (8m) discontinuity. The RMS error increases, but is still well within acceptable levels for most applications.

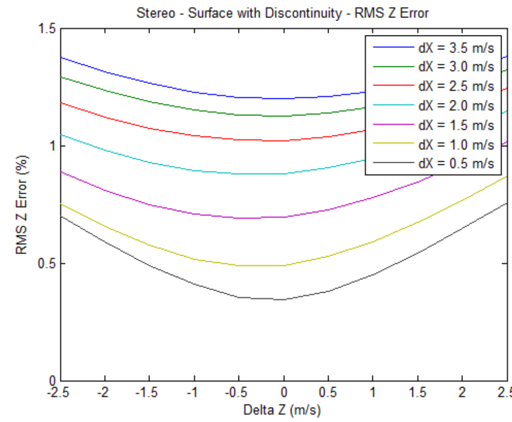

Figure 6

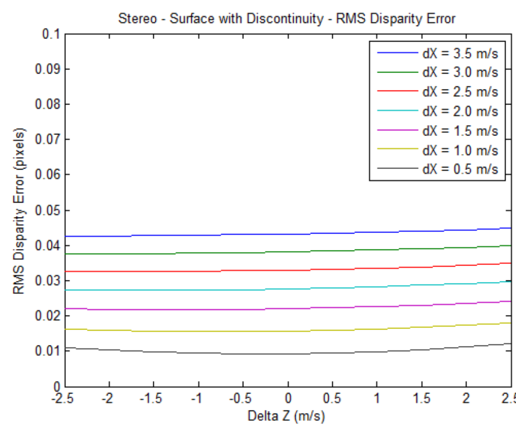

Figure 7

5.1.2 Coaxial Camera Rig

For the coaxial camera rig we determine the accuracy of the resulting image registration by reconstructing the depth map along radial lines as determined by the results of registration and compare the reconstructed depth map with the original scene geometry computing the RMS depth and disparity error.

For our synthetic flow images we used the same scene geometry as for the stereo camera rig. f_f = 4.8mm, f_b = 4.0mm, the camera has .002mm square pixels, velocity in the XY plane was varied from 0.5 m/s to 3.5 m/s and velocity along the Z-axis ranged from 2.5 m/s toward the camera to 2.5 m/s away from the camera. The camera frame rate was set to 30fps. We set $\gamma = 1 \cdot 10^{11}$ and $\alpha = 5 \cdot 10^{-5}$.

Figures 8 and 9 show the results for a smooth scene for a horizontal line. With the exception of the slowest XY displacement (0.5 m/s) and highest Z displacements, RMS depth error is < 0.15%. The shape of the curves suggest that there may be limitation on how large the Z displacement can be relative to the camera geometry and the XY displacement and still produce good results. We believe that this limitation may be due to cancellation which can occur between optical flow produced by lateral translation and the flow produced by forward translation in certain areas within the image.

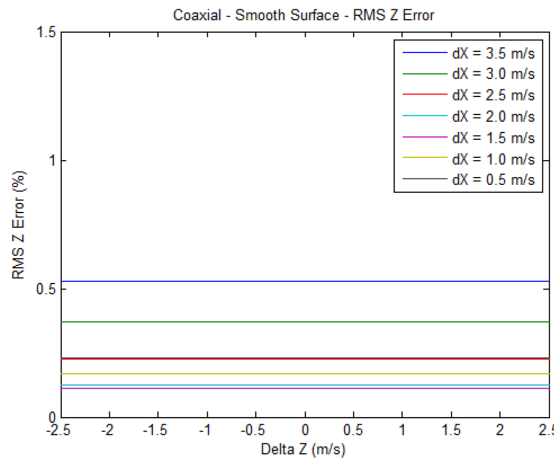

Figure 8

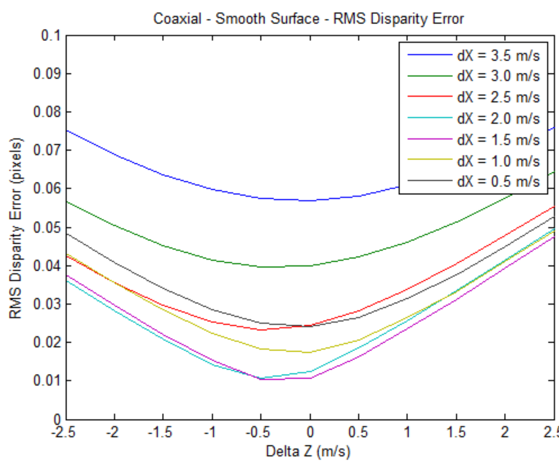

Figure 9

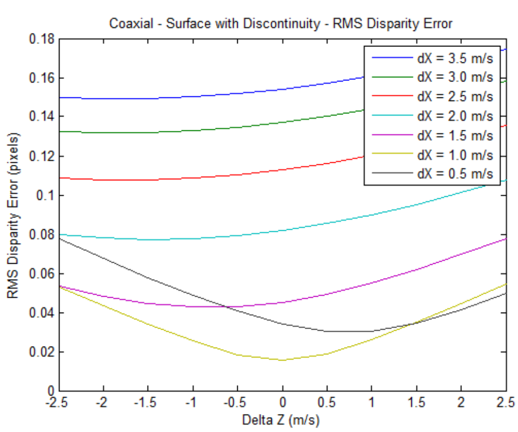

Figure 10

Figures 10 and 11 show the results for a scene with a large (8 m) discontinuity. The RMS error increases, but is still within acceptable levels for a wide range of applications.

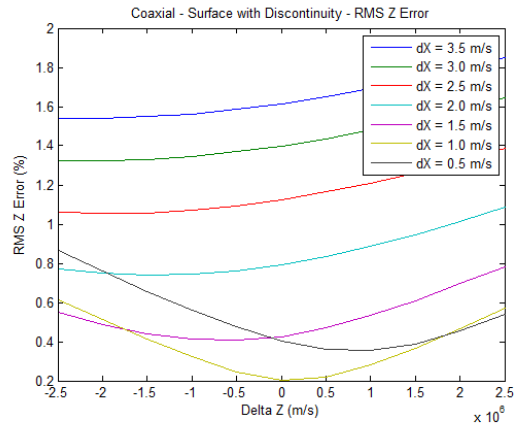

Figure 11

5.2. Real Flow Fields

5.2.1 Stereo Camera Rig

The multi-modal stereo camera consists of one camera with an RGB sensor and a second camera with an IR sensor. The camera rig was mounted on a precision XY table and the camera rig was translated a known distance between frames. Accuracy was determined by comparing the estimated camera rig displacement to the known camera rig displacement and converting to disparity.

For the multimodal stereo camera, our scene is shown in figure 12. There are small occlusions between the geometric shapes. Velocity in the XY plane was varied between 0.15 and 0.3 m/s, which when scaled to match our synthetic images would be about 4 m/s. The cameras in the stereo rig had 5.3 micron (IR) and 6 micron (RGB) square pixels and 7.0 mm (IR) and 7.7 mm (RGB) focal lengths. The images were corrected for the difference in pixel size and focal length. The baseline b = 75 mm. Gamma ranged from .2 to .5 and alpha was set at 0.01. To compute optical flow, we used the large scale optical flow algorithm from Brox et al. [9].

Figure 13 shows the results of our tests. At higher velocities the RMS disparity errors ranged from under 3 pixels to slightly over 8 pixels. As the velocity drops the disparity error increases. We believe this is due to the errors in optical flow being higher as a percent of the flow for flow fields with smaller magnitudes. The results at higher lateral velocities compare vary favorably to existing multi-modal camera registration techniques.

Figure 12

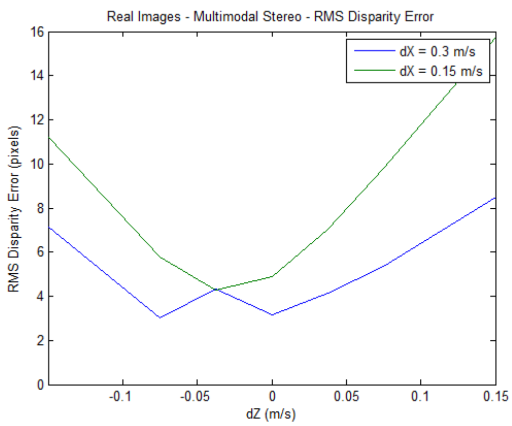

Figure 13

5.2.2 Coaxial Camera Rig

The coaxial camera rig consists of a pair of cameras with RGB sensors on the same XY table described above. The camera arrangement is shown in Figure 14. Coaxial camera depth accuracy was also determined by estimating the camera movement using the estimated depth map and optical flow field and comparing the estimated camera rig displacement to the actual displacement..

Our scene (Figure 15) consisted of a 10 cm diameter by 17 cm tall cylinder located 75 cm from the optical center of the front camera in the camera rig and a planar background located 115 cm from the optical center of the front camera. There is a relatively large discontinuity between the cylinder and the planar background similar in scale to that of our second set of synthetic experiments. Velocity in the XY plane was 0.3 m/s, which when scaled to match our synthetic images would be 4 m/s. The cameras in the coaxial rig have 0.006 mm square pixels, focal lengths of 7.7 mm and 5.8 mm (front and back respectively), and b = 143.3 mm. We set $\gamma = 2 \cdot 10^6$ and $\alpha = .05$. As with the stereo rig, we used the large scale optical flow algorithm from Brox et al. [9].

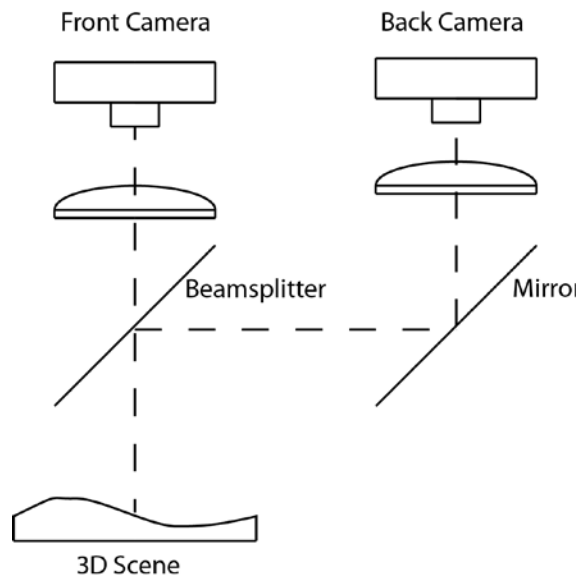

Figure 14

Figure 16 shows the results. The RMS disparity error is typically less than 1% except where we get the cancellation in the flow fields between the forward translation and the lateral translation. While substantially better than the results from the stereo rig, a coaxial camera rig requires considerably smaller disparity errors to produce the same depth errors as a stereo rig. However, for applications where image alignment is the objective, these results suggest that a coaxial camera rig is superior to a stereo rig.

Figure 15

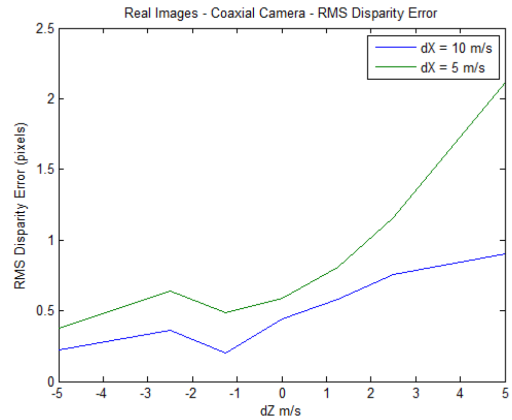

Figure 16

6. Discussion

Our results provide solid evidence that it's possible to register images using the optical flow fields provided that there is sufficient motion between the camera and the scene and that the scene has sufficient texture to produce optical flow. One advantages of our method is that images that don't have common pixel intensities or features can be aligned. Another advantage is that highly accurate sub-pixel alignment is possible in the center region of a coaxial camera. In both case this allows the estimation of dense disparity maps which can be converted into dense depth maps for 3D reconstruction and relative scene/camera rig velocity estimation.

With sufficient motion between the cameras and the scene and a scene that produces sufficient optical flow, our technique produces image alignment for a multimodal camera rig which is comparable to feature and pixel intensity based methods that align pairs of visible light images.

Our technique appears to be robust to flow fields that are not a good representation of the motion field as long as the flow fields in the two cameras reflect the same errors (e.g. the aperture problem and variation in illumination). This suggests that the intra camera images might be used as an additional term in the optical flow computation (e.g. intra-camera image smoothing) to improve both the optical flow computation and the results intra-camera image alignment.

Our results suggest that our technique could produce good results on moving multimodal camera rig (scanning security camera or vehicle mounted camera) and for a coaxial camera rig, allow stereo reconstruction in situations where a standard stereo baseline isn't feasible (e.g. endoscope or bore-scope).

References

[1] N. Asada, m. Baba, and A. Oda, "Depth from Blur by Zooming," in *Proceedings of the Vision Interface Annual Conference*, Ottawa, Canada, 2001.

[2] M. Baba, N. Asada, and T. Migita, "A Thin Lens Based Camera Model for Depth Estimation from Defocus and Translation by zooming," in *Proc. 15th International Conference on Vision Interface*, Calgary, Canada, 2002.

[3] F. B. Campo, R. L. Ruiz, and A. D. Sappa, "Multimodal Stereo Vision System: 3D Data Extraction and Algorithm Evaluation," *IEEE Journal of Selected Topics In Signal Processing, vol. 6, no.5*, 2012.

[4] G. Egnal, "Mutual information as a Stereo Correspondence Measure," *Technical Report MS-CIS-00-20, Computer and Information Science, University of Pennsylvania, Philadelphia, USA*, 2000.

[5] C. Fookes, A. Lamanna, and M. Bennamoun, "A new stereo image matching technique using mutual information," in *In Proceedings of the International Conference on Computer, Graphics and Imaging, CGIM'01, pages 168-173, Honolulu, USA, 2001. Iasted, ISBN 0-88986-303-2.*, 2001.

[6] C. Fookes, S. Maeder, S. Sridharan, and J. Cook, "Multi-Spectral Stereo Image Matching using Mutual Information," in *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, 2004.

[7] H. Gao, J. Liu, Y. Yu, and Y. Li, "Distance measurement of zooming image for a mobile robot," *International Journal of Control, Automation and Systems*, vol. 11, pp. 782-789, 2013.

[8] A. A. Goshtasby, *Image registration principles tools methods*: Springer, 2012.

[9] R. Hartly and A. Zisserman, *Multiple View Geometry in computer vision*: Cambridge University Press, 2003.

[10] R. Kirby, "Three Dimensional Surface Mapping System Using Optical Flow US2013321790A1," USA Patent, 2012.

[11] S. Krotosky and T. Mohan, "Registration of Multimodal Stereo Images using Disparity Voting from Correspondence Windows," in *IEEE Conf. on Advanced Video and Signal based Surveillance (AVSS'06)*, 2006.

[12] S. Krotosky and M. Trivedi, "Multimodal Stereo Image Registration for Pedestrian Detection," in *in Proc. IEEE Intell. Transp. Syst. Conf., Sep. 2006, pp. 109–114*, 2006.

[13] S. Krotosky and M. Trivedi, "Mutual information based registration of multimodal stereo videos for person tracking," *Computer Vision and Image Understanding*, vol. 106, pp. 270-287, 2007.

[14] J. Lavest, G. Rives, and M. Dhome, "Three Dimensional Reconstruction by Zooming," *IEEE Transactions on Robotics and Automation*, vol. 9, pp. 196-207, 1993.

[15] J. Lavest, G. Reves, and M. Dhome, "Modeling an Object of Revolution by Zooming," *IEEE Transactions*

on Robotics and Automation, vol. VOL. II, NO. 2, April 1995, 1995.

[16] J. Ma and S. I. Olsen, "Depth from Zooming," *J. Opt. Soc. Am. A* vol. 7, October 1990 1990.

[17] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two frame stereo correspondence algorithms," *International Journal of Computer Vision* vol. 47, pp. 7-42, 2002.

[18] R. Szeliski, *Computer Vision. Algorithms and Applications*. New York: Springer, 2011.

[19] J. M. Thomas Brox, "Large Displacement Optical Flow Desriptor Matching in Variational Motion Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2010.

[20] A. Toraby and G. Bilodeau, "Local self-similarity as a dense stereo correspondence measure for themal visible video registration," *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on,* pp. 61 - 67, 2011.

[21] A. Verri and T. Poggio, "Motion Field and Optical Flow: Qualitative Properties," *IEEE Trans. Pattern Analysis and Machine Intelligence, 11(5), pp. 490{498, May*, 1989.

[22] P. Viola and W. M. Wells, "Alignment by Maximization of Mutual Information," *Intl. J. of Computer Vision, vol. 24, no. 2, pp. 137–154,,* 1997.

[23] M. Yaman and S. Kalkan, "An iterative adaptive multi-modal stereo-vision method using mutual information," *Journal of Visual Communication and Image Representation,* vol. 26, pp. 115-131, 2015.

[24] Y. Zhang and K. Qi, "Snake-Search Algorithm for Stereo Vision Reconstruction via Monocular System," presented at the The 5th Annual IEEE Conference on Cyber Technology in Automation, and Control, Intelligent Systems, Shenyang, China, 2015.

[25] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing,* vol. 21, pp. 977-1000, 2003.