MULTI-VIEW IMAGE REGISTRATION FOR WIDE-BASELINE VISUAL SENSOR NETWORKS

Gulcin Caner[a], A. Murat Tekalp[b], Gaurav Sharma[a], Wendi Heinzelman[a]

[a] Electrical and Computer Engineering Dept., University of Rochester, Rochester, NY, 14627-0126
[b] College of Engineering, Koc University, Istanbul, Turkey
mtekalp@ku.edu.tr,{caner,gsharma,wheinzel}@ece.rochester.edu

ABSTRACT

We present a new dense multi-view registration technique for wide-baseline video/images that integrates a parametric optical flow-based approach with a sparse set of feature correspondences, based on a locally planar approximation of a nonplanar scene. The proposed method can deal with illuminance variations between the views, which is critically important for wide-baseline applications. It differs from existing work on wide-baseline image registration in that it requires only image information and provides dense matching without computing any camera calibration matrices or performing any prior scene segmentation. These characteristics render the method suitable for practical deployment in visual sensor networks, towards which the current work is directed. We demonstrate the performance of the proposed method on simulated multi-view images of a virtual 3D world composed of piece-wise smooth textured surfaces, as well as real wide-baseline images of nonplanar textured surfaces.

Index Terms— Local image registration, wide-baseline, Wiener-based affine model estimation

1. INTRODUCTION

Multi-view image registration is an important step in many computer vision and video processing applications such as camera calibration, 3D scene reconstruction and creation of panaromic views, etc. Numerous methods have been proposed to solve this problem for different ranges of baselines (i.e., separations) between the cameras.

In the literature, there has been more emphasis on small-baseline applications. Multi-view image registration techniques for the case of small-baseline usually assume that a single homography (projective transformation) or lower order parametric models such as pseudo-perspective and affine, can effectively model the spatial transformation between the multiple views. These techniques [1] [2] [3] [4], which are classified as either intensity-based or feature-based methods, become insufficient to solve the problem of wide-baseline image registration due to several critical issues that arise with wide-baseline, including non-negligible parallax, depth discontinuities, and occlusions.

In wide-baseline scenarios, the focus has been on the establishment of a set of view-invariant feature correspondences across the multiple views for the end goal of camera calibration and 3D scene reconstruction [5] [6]. Most of these techniques aim to first compute camera matrices (i.e., camera external and internal parameters) from feature correspondences and then provide the dense matching of multi-view images using the calibration information. In a recent work [7], an initial set of feature correspondences is expanded for dense matching of rectified wide-baseline images. In addition, use of prior scene data in the form of an image sequence has also been proposed to register uncalibrated wide-baseline images [8] [9].

In this paper, we propose a new multi-view image registration method for wide-baseline applications. The main contributions of this paper are: *i*) we employ a locally planarized scene model; hence, a locally-varying affine disparity field model; *ii*) we develop a locally-varying affine parameter estimation technique, which integrates optical flow-based methods with sparse feature correspondences, where the feature correspondences enable coarse-level registration, and the locally-varying affine model provides fine-level dense registration; and *iii*) the method can deal with illuminance variations between the views, which is critical for wide-baseline applications. Another feature of the proposed method is that it is computationally simpler and more robust than existing wide-baseline dense matching techniques, since it does not require pre-processing of the input images for camera calibration or scene segmentation. These features make the method well-suited for practical use in visual sensor networks.

2. THEORY OF THE PROPOSED METHOD

This section first presents the scene and disparity field models, which form the basis of the proposed locally varying affine registration method; then, the proposed algorithm to compute the model parameters recursively at each pixel is presented.

2.1. Locally Planar Scene Model

We assume a locally planarized scene model, which is analogous to linearization of a non-linear function. As a result, the 2D displacement/disparity field between the corresponding planes in the multiple views can be modeled as a homography. This can be further approximated by a simpler locally varying affine model [5], that is computationally advantageous due to fewer number of parameters.

To overcome the aperture problem, the affine model parameters at each pixel are estimated over a block of pixels. Naturally, the size of the block at a given pixel relates to the size of the planar patch that is tangent to the actual surface for that location. Small block sizes are required to track variations in the surface normal, while large block sizes are required to ensure consistency of the affine parameter estimates. Experimental results indicate that the parameter estimation method developed in the following provides successful results with block sizes as small as 3 x 3; hence the ability to track surface nonlinearities. Note that our scene model is different from the common approach, where the input scene is modeled as a union of a small number of planes, which typically requires scene segmentation before plane-by-plane image registration [8].

This work is partly supported by the National Science Foundation under grant number ECS-0428157.

2.2. Computation of Locally Varying Affine Model Parameters

The well-known optical flow constraint uses the linearized (first-order) Taylor Series expansion, where the range of displacement to be computed is limited. However, when the displacements among the two images are large, this constraint cannot be used to compute the displacements accurately, unless a multi-scale (pyramid) approach is followed. In order to compute larger displacements without the computational complexity of the multi-scale approaches, we follow the Wiener-estimation based solution [10]. We extend the translational motion model presented therein to an affine motion model and also compensate for the local illumination changes by subtracting the local means. In the following, we give the recursive formulations for the locally varying affine model, where the displaced frame difference, $dfd()$, at a location, X_c is defined as follows:

$$dfd(X_c, A(X_c), t(X_c)) = \\ (I_s(A(X_c)X_c + t(X_c)) - \mu_s) - (I_c(X_c) - \mu_c) \quad (1)$$

where I_c and I_s denote the current and the source images, i.e., the motion field is estimated from the current image towards the source image. μ_c and μ_s denote the local (gray-level) mean at X_c in the current image and at its corresponding location in the source image, respectively. $A(X_c)$ and $t(X_c)$ are rotation/scale/skew matrix and translational vector of the affine model, respectively, at the pixel location, X_c, where

$$A(X_c) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad and \qquad t(X_c) = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

For concise notation, we abbreviate $A(X_c)$ and $t(X_c)$ as A and t, respectively. Under the assumption that the affine model parameters at the current pixel are obtained as updates of parameters from the previous pixel, we can rewrite Eqn. (1) as:

$$dfd(X_c, A, t) = \\ (I_s((A^i + a^i)X_c + (t^i + u^i)) - \mu_s) - (I_c(X_c) - \mu_c) \quad (2)$$

where A^i and t^i are the initial affine model parameters at location X_c, and a^i and u^i are the corresponding updates on A^i and t^i, respectively. Expanding the the first term on the right side of Eqn. (2) into a Taylor series about $(A^i X_c + t^i)$, we obtain:

$$dfd(X_c, A, t) = \\ (I_s(A^i X_c + t^i) + (a^i X_c + u^i)\nabla I_s(A^i X_c + t^i) \\ + h.o.t. - \mu_s) - (I_c(X_c) - \mu_c) \quad (3)$$

where 'h.o.t.' denotes the higher order terms in the Taylor series expansion, i.e., linearization error. By denoting the displaced frame difference due to the i^{th} iteration of the affine model, at location X_c by $dfd(X_c, A^i, t^i)$, Eqn. (3) can be written as:

$$dfd(X_c, A, t) = \\ dfd(X_c, A^i, t^i) - \mu_s + \mu_c \\ + (a^i X_c + u^i)\nabla I_s(A^i X_c + t^i) + h.o.t. \quad (4)$$

In order to compute the update parameters, (a^i, u^i) in Eqn. (4), we impose a uniform affine motion model constraint over a block, B, of pixels around the current pixel, X_c, and set $dfd(X_c, A, t)$ to zero within that neighborhood. This results in a set of equations:

$$-(dfd(X(j), A^i, t^i) - \mu_s + \mu_c) = \\ (a^i X(j) + u^i)\nabla I_s(A^i X(j) + t^i) + h.o.t., \forall X(j) \in B \quad (5)$$

where $X(j) = [x_1(j), x_2(j)]$ denotes location of the j^{th} pixel in block, B. Grouping these into matrix/vector format yields

$$z = G.upd(X_c) + n \quad (6)$$

where n is the vector of linearization error terms, z and G are defined as in Eqn. (8) and $upd(X_c) = [a_{11}, a_{12}, a_{21}, a_{22}, u_1, u_2]$ is the vector of update terms on the affine model parameters. Making some simplifying assumptions on the covariance matrices of $upd(X_c)$ and n (see [10, Chap. 7]), the minimum mean square solution of Eqn. (6) is

$$upd(X_c) = (G^T G + \mu I)^{-1} G^T z \quad (7)$$

where I is the identity matrix, and μ is a damping factor.

2.3. Combining Optical Flow-based Affine Parameter Estimation with Feature Correspondences

Wide-baseline imaging scenarios result in greater disparity between the images, which optical flow-based approaches become inadequate to handle without appropriate initialization [8]. To solve this problem and still have the merits of gradient-based approaches (i.e., dense matching and sub-pixel precision in the matching accuracy), we utilize a feature-based correspondence estimation method [11]. We integrate a set of sparse feature correspondences with the optical flow-based approach presented in Section 2.2, such that appropriate initialization, (A^i,t^i), for the local affine model parameters is provided by the feature correspondences.

Figure 1 shows a flow diagram of the integration of feature correspondences with the optical flow-based affine parameter estimation. We commence the scan of the image from the feature point closest to the top-left corner. For the first pixel in the image scan order, (A^i, t^i) are initialized using this feature correspondence by setting t^i to the displacement between the correspondent feature points, and A^i to the identity matrix. Locally varying affine parameters at each pixel on the scan order are then updated recursively using Eqn. (7). Appropriate re-initialization with the feature correspondences is performed when required, as shown in Figure 1. The complete image region is scanned by beginning with this first corresponding feature point and proceeding in the first image along a Hilbert curve [12] in either direction (toward top-left and bottom-right corners, respectively). The corresponding locations in the second image are determined by transforming these using the current affine transformation estimates. The use of the Hilbert curves helps better preserve contiguity and locality [13].

3. EXPERIMENTAL RESULTS

We apply the proposed wide-baseline image registration technique to register multi-view images captured in two different scenarios, *i*) in a virtual 3D world and *ii*) in a real world. In the implementation of the proposed method, we set the damping factor (i.e., μ) to 10 in the minimum mean square solution. In both experiments, the best block size for the uniform affine model constraint has been determined by search over a discrete set of candidate block sizes.

In the simulation experiment, we utilize a 3D object composed of multiple piece-wise smooth planar surfaces, as shown in Figure 2. Note that we have chosen textured surfaces to ensure that local variations in registration can be estimated reliably. In the absence of such texture, correspondences cannot be unambiguously determined (generalized aperture problem). Figures 3 and 4 show the multi-view images captured from two different viewing angles, where the camera motion is composed of 3D rotation and translation. Plane

$$z = \begin{pmatrix} -(dfd(X(1), A^i, t^i) - \mu_s + \mu_c) \\ -(dfd(X(2), A^i, t^i) - \mu_s + \mu_c) \\ \vdots \end{pmatrix}, G = \begin{pmatrix} x_1(1)\frac{\partial I_s(A^i X(1)+t^i)}{\partial x_1} & x_2(1)\frac{\partial I_s()}{\partial x_1} & x_1(1)\frac{\partial I_s()}{\partial x_2} & x_2(1)\frac{\partial I_s()}{\partial x_2} & \frac{\partial I_s()}{\partial x_1} & \frac{\partial I_s()}{\partial x_2} \\ x_1(2)\frac{\partial I_s(A^i X(2)+t^i)}{\partial x_1} & x_2(2)\frac{\partial I_s()}{\partial x_1} & x_1(2)\frac{\partial I_s()}{\partial x_2} & x_2(2)\frac{\partial I_s()}{\partial x_2} & \frac{\partial I_s()}{\partial x_1} & \frac{\partial I_s()}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \quad (8)$$

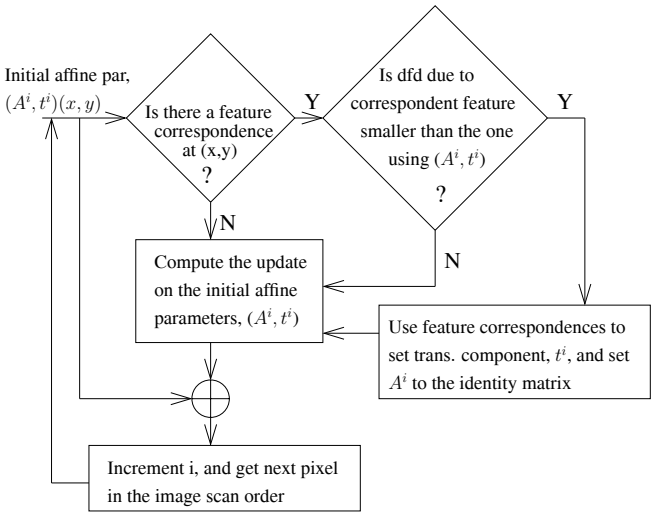


Fig. 1. Image registration combining affine parameter estimation with feature correspondences. Affine model parameters are estimated recursively for each pixel on the image scan order.

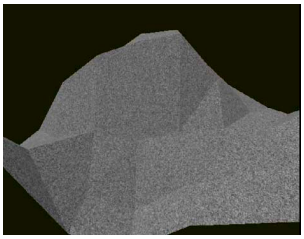


Fig. 2. Virtual 3D world.

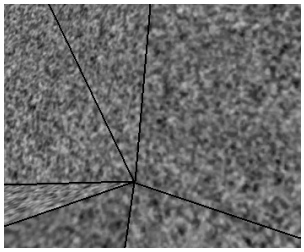


Fig. 3. Simulated view 1 (Source image).

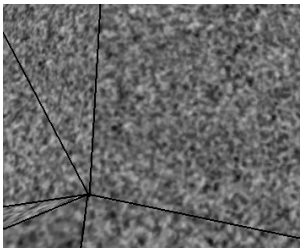


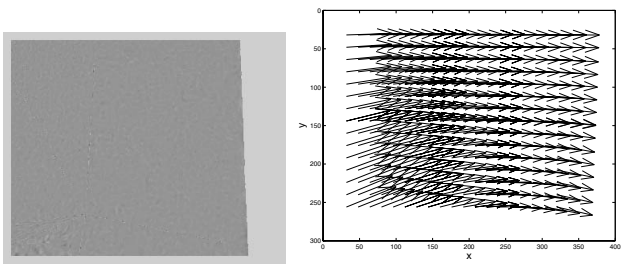
Fig. 4. Simulated view 2 (Current image).

boundaries are drawn on Figures 3 and 4 for clarification. We apply the proposed registration method by imposing the uniform affine model constraint over a block of size (11x11) throughout the image. Figures 5-(a) shows the registration error image. In addition to the high pSNR (37.75 dB) of the resulting error image, the estimated motion field can be observed to be consistently varying, in Figure 5-(b).

For the real experiment, we apply the proposed registration method to two wide-baseline images captured from a non-planar surface of a couch by a digital camera. Figures 7-(a) and 7-(b) show the source and current images, respectively, where the motion field is estimated from the current image towards the source image. We apply the proposed registration method by imposing the uniform affine model constraint over a block of size (3x3) throughout the image. We present the preliminary results for the registration error image, and the estimated motion field in Figure 7-(c) and Figure 6, respectively. The pSNR value of the registration error image is computed to be 33.44 dB. From Figure 6, the estimated motion field can be observed to be consistently varying everywhere, except the locations where an accurate registration cannot be achieved due to lack of adequate texture.

4. CONCLUSION

We propose a new method for wide-baseline image registration, by combining optical flow-based motion estimation with feature correspondences. We model the 3D world locally as planar and apply a recursive algorithm to update the affine model parameters for each planar patch on a local basis. Feature correspondences are utilized to initialize the affine model in order to track the large distortions in the wide-baseline imaging scenarios. The method has advantages over other wide-baseline registration techniques, such that it does not require any pre-processing of the scene or the cameras (i.e., scene segmentation and/or camera calibration). Experimental results show that by choosing an appropriate block size for the uniform motion model constraint and suitably integrating the feature correspondences, wide-baseline images can be registered with high accuracy. Both the registration error image and the estimated motion field validate the performance of the method. Future work includes using a spatially adaptive window size for the uniform affine model constraint and increasing robustness of the technique to the possible



(a) (b)

Fig. 5. Virtual 3D world **(a)** Error image after registration; **(b)** Estimated motion field using the proposed method.

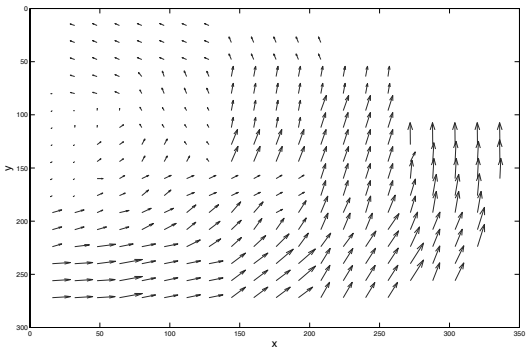


Fig. 6. Real nonplanar scene: Estimated motion field (scaled) using the proposed method.

occlusions in the scene.

5. REFERENCES

[1] Y. Caspi and M. Irani, "Spatio-temporal alignment of sequences," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 24, no. 11, pp. 1409–1424, 2002.

[2] L. Lee, R. Romano, and G. Stein, "Monitoring activities from multiple video streams:establishing a common coordinate frame," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 22, no. 8, pp. 758–767, 2000.

[3] P.H.S. Torr and A. Zisserman, "Feature based methods for structure and motion estimation," in *Proc. Int. Workshop Vision Algorithms*, 1999, pp. 278–295.

[4] H. Liu, R. Chellappa, and A. Rosenfeld, "Accurate dense optical flow estimation using adaptive structure tensors and a parametric model," *IEEE Trans. Image Proc.*, vol. 12, no. 10, pp. 1170–1180, 2003.

[5] A. Baumberg, "Reliable feature matching across widely separated views," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, 2000.

[6] C. Strecha, T. Tuytelaars, and L. V. Gool, "Dense matching of multiple wide-baseline views," in *IEEE Intl. Conf. Comp. Vision.*, 2003.

[7] Z. Megyesi and D. Chetverikov, "Affine propagation for surface reconstruction in wide baseline stereo," in *IEEE Intl. Conf. on Pattern Recog.*, 2004.

[8] J. Xiao, Y. Zhang, and M. Shah, "Adaptive region-based video registration," in *IEEE Workshop on Motion*, Jan. 2005.

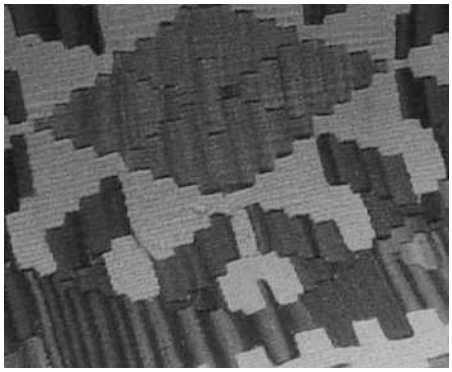
[9] A. Roy-Chowdhury and R. Chellappa, "Wide baseline image registration with application to 3-d face modeling," *IEEE Trans. Multimedia*, vol. 6, no. 3, pp. 423–434, June 2004.

[10] A. M. Tekalp, Ed., *Digital Video Processing*, Prentice Hall, Upper Saddle River, NJ, 1995.

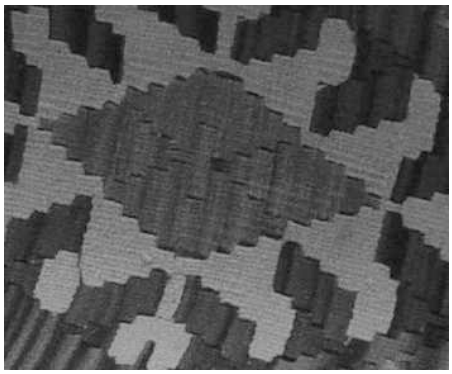
[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intl. J. Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] H. Sagan, *Space-filling curves*, Springer, Berlin, 1994.

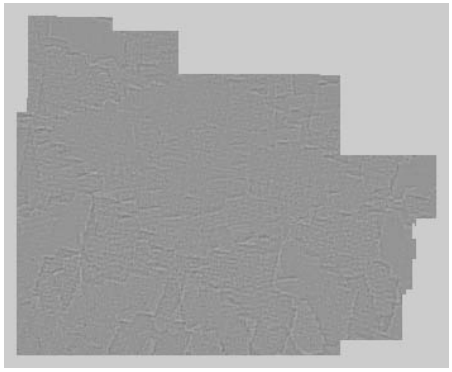
[13] G. Caner, A. M. Tekalp, G. Sharma, and W. Heinzelman, "Local image registration by adaptive filtering," *IEEE Trans. Image Proc.*, submitted August 2005. Accepted for publication.



(a)



(b)



(c)

Fig. 7. Real nonplanar scene **(a)** View 1 (Source image); **(b)** View 2 (Current image); **(c)** Error image after registration using the proposed method.