Analyzing spatial dependencies on the reliability of 2D to 3D flow estimation

Subhajit Chaudhury[1,a)] Gaku Nakano[1,b)]

1. Introduction

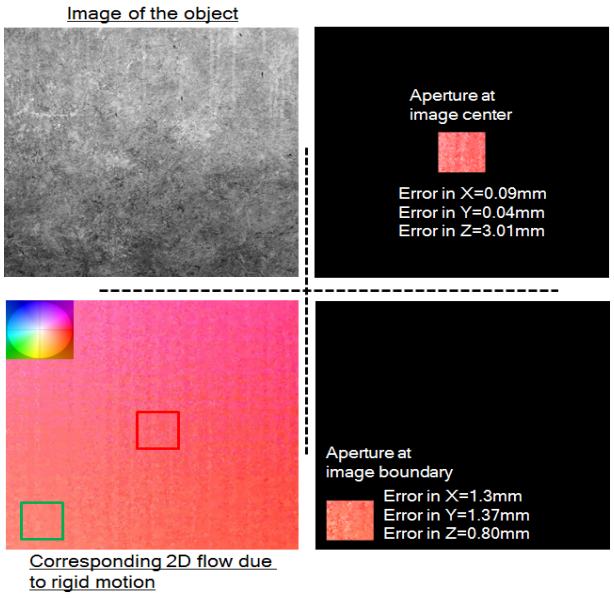
Recovering pixel-wise 3D deformation from monocular images is a highly under-constrained problem. This problem is similar to optical flow estimation which is also an under-constrained problem. Inspired from the similarity with optical flow techniques[3], [7], [9], we solve the dense 3D non-rigid reconstruction problem using a variational approach which directly estimates dense 3D flow from dense 2D correspondences between subsequent frames.

The major contribution of this paper is to present a novel aperture problem related to converting 2D optical flow to 3d flow. Conventional techniques assume that the reliability of 3D motion estimation from image sequences is independent of the spatial location of estimation. However in this paper, we show that estimation of 3D motion parameters are dependent on the spatial location of estimation. The proposed aperture problem tells that, from 2D dense motion, 3D motion along optical axis can be computed with comparatively higher confidence on image boundary while with low confidence at image center. Similarly motion perpendicular to the optical axis can be found with high confidence at image center and with comparatively lower confidence at the image boundary. This phenomenon is shown in figure 1, where we can see the error in X,Y,Z estimation at the image center and boundary concurs with our hypothesis. We perform synthetic experiments to demonstrate our hypothesis. Furthermore, we employ a weighted approach to successfully find 3D non-rigid deformation from 2D correspondences for real images.

The detailed pipeline for non-rigid 2D to 3D flow estimation is given next. Given two monocular images, our method provides a pipeline for dense non-rigid 3D flow estimation directly from dense 2D optical flow between two images. We assume that the rigid configuration of the initial frame is provided beforehand. We minimize the re-projection error between the observed 2D motion and projected inter-frame 3D motion assuming local rigidity. We assume a calibrated camera with perspective projection model, which enables us to find highly non-rigid motion along the camera optical axis also.

[1] NEC Central Research labs, Kanagawa-ken, Nakahara-ku, 1753, Shimonumabe, 211-8666 Japan
[a)] s-chaudhury@ap.jp.nec.com
[b)] g-nakano@cq.jp.nec.com

Our method is closest to the work on shape from template[2], [11], [12], [13]. These methods include creating a 3D mesh configuration and assigning pixels to each vertices of the mesh and imposing additional geometric constraints to recover the 3D shape in the target frame. Although our method is most closely related to these kind of methods, our analysis of the reliability of obtaining 3D flow from 2D flow based on spatial location is what that makes our paper different. Also we recover 3D deformation directly from 2D correspondences and hence provide a much simpler implementation framework which is easily scalable for multi-frame sequences.



(a)

Fig. 1: (Top-left) Showing an image of virtual plane texture mapped with concrete material texture; (Bottom-left) Corresponding 2D flow; (Right) At the image center Z-motion estimation is less reliable and X,Y movement is more reliable. On the image boundaries, Z-motion estimation is more reliable than X,Y estimation. This is shown by the errors in estimation.

2. Spatial Dependencies on Reliability of 2D to 3D flow

In optical flow, aperture problem refers to the inability to discern optical flow value in the direction perpendicular to

the local image gradient when the viewing area is limited to a small neighborhood. In this paper we extend the definition of conventional "aperture problem", for optical flow, to a more general setting and present a similar problem for inferring 3d flow from local 2D optical flow. Following a similar approach to the work by Lucas and Kanade ([9],[10]) for the 2D flow problem, we assume that for every pixel, 3D flow is constant in a local neighborhood around it. With this local rigidity assumption around a pixel p, we want to minimize the error between observed 2D motion vector and projected 2D motion from 3D flow onto the image plane by perspective projection. This is mathematically represented as,

$$\min_{(\delta x_p, \delta y_p, \delta z_p)} \sum_{i \in \Omega_p} \left\| \underbrace{\frac{1}{z_i} \begin{bmatrix} f & 0 & u_i \\ 0 & f & v_i \end{bmatrix}}_{A_i} \begin{bmatrix} \delta x_p \\ \delta y_p \\ \delta z_p \end{bmatrix} - \begin{bmatrix} m_{x_i} \\ m_{y_i} \end{bmatrix} \right\|^2 \quad (1)$$

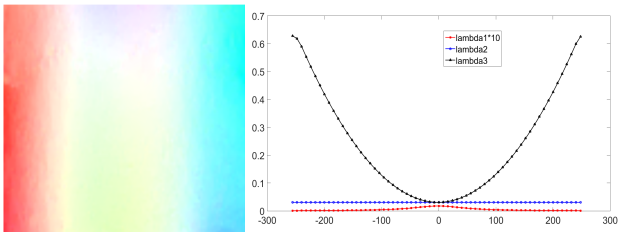
where we assume, for all pixels in the neighborhood of pixel $p(u_p, v_p)$ are the zeros centered 2D image co-ordinates,the 3D flow at pixel p is $[\delta x_p, \delta y_p, \delta z_p]^T$, f is the focal length of the camera, z_i is the depth in the initial frame from the camera and (m_{x_i}, m_{y_i}) is the 2D motion at the pixel p.

In 2D optical flow, corner points(image gradients form a solvable set of linear equations) is considered as salient points where optical flow can be determined with high confidence. However unlike optical flow where the aperture problem depends on the image gradients only and is independent of spatial location on the image, we show that for 3D flow estimation, aperture problem is a fundamental phenomenon that depends on the spatial location of the image co-ordinate system.

In order to explain this effect, let us define the motion basis at each pixel for X, Y, Z motion. The relation between real world 3D motion and the projected 2D motion on the image plane at the p^{th} pixel is given as,

$$\frac{1}{z_p} \begin{bmatrix} f & 0 & u_p \\ 0 & f & v_p \end{bmatrix} \begin{bmatrix} \delta x_p \\ \delta y_p \\ \delta z_p \end{bmatrix} = \begin{bmatrix} m_{x_p} \\ m_{y_p} \end{bmatrix} \quad (2)$$

where f is the focal length of the camera, (u_p, v_p) are zeros centered image co-ordinates, that is $u_p = u - u_0, v_p = v - v_0$, where (u_0, v_0) are the optical center of the camera expressed in pixels. From equation 2, we observe that there is a dependence on the initial depth(z_p) at the pixel p, which requires an initial depth map for the first frame of the sequence. We refer to the motion along optical axis of the camera as Z-motion and the motion on a plane perpendicular to the optical axis is referred to as X-Y motion. From equation 2 we can derive the basis motion for 3D motion contribution on 2D flow. The motion basis vectors at a pixel location $p = (u_p, v_p)$ are $\mathbf{b}_x = f\hat{i}, \mathbf{b}_y = f\hat{j}, \mathbf{b}_z = u_p\hat{i} + v_p\hat{j}$, where $(\hat{i}, \hat{j})$ are the orthonormal basis of the image co-ordinate system. The task of obtaining 3D motion from 2D flow is equivalent to finding the projection of observed 2D motion vector into these motion basis vectors $\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z$. It is evident that this projection can have infinite solutions and thus additional constraints are required to estimate the 3D flow for each pixel. However from the local rigidity assumption, we obtain a set of linear equations given by equation 1 that can be solved, if the matrix of the right hand side is well conditioned.



(a) (b)

Fig. 2: (a) The 2D correspondence map (Middlebury[1] optical flow color mapped) for the paper bending sequence in [11] (b) Plot of eigenvalues for the matrix from equation 1 to find δx_p, δy_p, δz_p .

The contribution to 2D motion due to Z-motion vanishes at the optical center (where $(u_p, v_p) = (0, 0)$) and hence the local solution(solution of equation 1) is ill-conditioned at the optical center. Hence we cannot estimate the component of 3D motion along the optical axis. Figure 2(a) gives the 2D motion vector visualization due to the paper bending sequence from [11], which shows that 2D motion vanishes at the optical center. However at the image boundary, contribution due to Z-motion dominates that due to X-Y motion on 2D motion. Thus Z-motion can be estimated with high confidence at the image boundary while X-Y motion estimation is prone to motion errors. This fact can be confirmed from the eigenvalue(for the matrix $\sum A_i^T A_i$) plot of figure 2(b), which shows that at the optical center two eigenvalues have almost same magnitude while only one eigenvalue has large value (corresponding to reliable estimation of Z-motion) at the image boundary. However at the image boundary, there will be comparatively more error in Z-motion estimation if there is presence of considerable 3D lateral movements.

3. Solving 3D flow using a local-global optimization approach

The above discussion suggests that there is different form of "aperture problem" for 3D motion estimation, which is dependent on spatial location of the image. We propose a simple 2D to 3D flow estimation and show that non-linear deformations with high Z-motion contribution at the image boundary can be estimated with higher precision compared to deformations which have major Z-motion contribution at the image center. We extend the local-global approach from [3] for finding 3D flow from 2D flow. The optimization makes use of a local evidence term and a regularization term for smoothness of the 3D reconstruction. Overall energy functional is the summation of the data term and

the regularization term, given as $E = E_d + \alpha E_s$, where the parameter α controls the strength of smoothness in the solution. We minimize this energy functional by standard iterative re-weighted least squares(IRLS) technique, similar to the optical flow computation in [8], which consists of two steps for every iteration : (1) Compute weights using the solution in the current iteration (2) Update the solution using the weights computed in the previous step(1) by solving a set of linear equations. The above two steps are repeated in every iteration, until convergence. We present stepwise details of our method in algorithm 1.

Algorithm 1 Algorithm for 2D to 3D flow computation

Input: Dense 2D correspondences $\mathbf{m}_x, \mathbf{m}_y$, Camera calibration matrix K, depth map of rigid template Z

Output: Dense 3D flow between reference and target frame $\delta\mathbf{x}, \delta\mathbf{y}, \delta\mathbf{z}$

Initialize: the 3D flow parameters $\delta\mathbf{x}^0, \delta\mathbf{y}^0, \delta\mathbf{z}^0$ to zero, residual 2D motion $\mathbf{e}_x = -\mathbf{m}_x, \mathbf{e}_y = -\mathbf{m}_y$

while *not converge or* $k \leq$ *maxiter* **do**

(1) (**Re-weighting**): Compute the generalized Laplacian $L = D_x^T \mathbf{\Phi} D_x + D_y^T \mathbf{\Phi} D_y$, the weights $\mathbf{\Psi}^k$ and the motion mapping matrix $\mathbf{J}^K$ for the current iteration.

(2) (**Local-global solution**): Compute incremental 3D flow update $(d\mathbf{U}^k, d\mathbf{V}^k, d\mathbf{W}^k)$ by solving the linear equation relating the incremental 3D flow with the residual error in projected 2D motion

$$\begin{bmatrix} \mathbf{\Psi}^k J_{1,1} + \alpha L & 0 & \mathbf{\Psi}^k J_{1,3} \\ 0 & \mathbf{\Psi}^k J_{2,2} + \alpha L & \mathbf{\Psi}^k J_{2,3} \\ \mathbf{\Psi}^k J_{1,3} & \mathbf{\Psi}^k J_{2,3} & \mathbf{\Psi}^k J_{3,3} + \alpha L \end{bmatrix} \begin{bmatrix} dU^k \\ dV^k \\ dW^k \end{bmatrix} = - \begin{bmatrix} \mathbf{\Psi}^k J_{1,4} + \alpha L \delta\mathbf{x}^k \\ \mathbf{\Psi}^k J_{2,4} + \alpha L \delta\mathbf{y}^k \\ \mathbf{\Psi}^k J_{3,4} + \alpha L \delta\mathbf{z}^k \end{bmatrix}$$

(3) **Update** the 3D flow

$$\begin{bmatrix} \delta\mathbf{x}^{k+1} \\ \delta\mathbf{y}^{k+1} \\ \delta\mathbf{z}^{k+1} \end{bmatrix} = \begin{bmatrix} \delta\mathbf{x}^k \\ \delta\mathbf{y}^k \\ \delta\mathbf{z}^k \end{bmatrix} + \begin{bmatrix} d\mathbf{U}^k \\ d\mathbf{V}^k \\ d\mathbf{W}^k \end{bmatrix}$$

end

We explain the steps of algorithm 1. It takes dense 2D motion between two frames as input and produces a 3D flow as output. This is an iterative process, where the in each step we compute weight matrix for the local-global solution(re-weighting) and solve it to obtain the dense 3D map update. This update is accumulated in each step to obtain the final solution.

4. Diffusion of 3D flow

Both the local estimation data term and regularization term contributes to improving the accuracy of 3D flow estimation. The data term based on local estimation provides robustness to motion noise. The regularization term on the other hand is responsible for diffusion of 3D flow from regions of high reliability of flow estimation to regions of lower reliability. Based on the discussion of aperture problem in section 2, this amounts to Z-motion estimation diffusing radially inwards from areas on the image boundary to the optical center while X-Y motion diffusing radially outwards from the optical center. This observation is supported by experimental evidence with Gaussian and 2D parabola shaped deformations as mentioned in the experimental evaluation section.

5. Experimental evaluation

In this section we perform experimental evaluation of our proposed hypothesis about aperture problem in 2D to 3D flow. Also we show our results on real deformations using our proposed 2D to 3D non-rigid estimation algorithm discussed in section 3.

5.1 Synthetic Rigid sequence

To verify the our hypothesis, we compute X,Y,Z motions at the image center and at the image boundary and test the errors in measurements at these spatial locations. We simulate an synthetic experiment using Computer Graphics software(OpenGL). For this purpose we chose a rigid planar movement of dX=10mm, dY=0mm and dZ=10mm, with focal length of 2268 pixels and distance of 10m. We compute the X,Y,Z motion at the image center and at the image boundary. This scenario is depicted in figure 1 where the chosen locations are shown. The errors in the estimation is shown in table 1. We used a block-size of 128×128 for finding local 3D motion estimates.

Table 1: Table showing error in X,Y,Z 3D motion estimation at image center and image boundary.

Image	Image center	Image boundary
Error in X	0.09	1.30
Error in Y	0.04	1.37
Error in Z	3.01	0.8

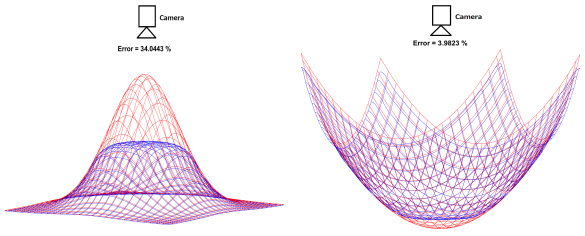
From Table 1, we see that the error for Z-estimation is more at the image center compared to that at the image boundary. For X,Y motion it is seen that errors at the image boundary is higher compared to the image center. This result produces verification of our hypothesis and demonstrates that reliability of 3D flow estimation depends on spatial location.

5.2 Synthetic Non-rigid sequence

In addition to the rigid 3D motion estimation experiment, we perform additional experiments for non-rigid deformation estimation. We choose two types of deformation, one where the Z-deformation at the image center is large compared to that at the boundary and second where the Z-deformation at the image boundary is more than that in the center.

We adopt the RMS error metric defined by [4] and define the percentage 3D error as $error_{3D} = \frac{||M_f - M_f^{GT}||_F}{||M_f^{GT}||_F}$, where M_f is the concatenated recovered 3D motion $(\delta x_p, \delta y_p, \delta z_p)$ and M_f^{GT} is the concatenated ground truth 3D motion $(\delta x_p^{GT}, \delta y_p^{GT}, \delta z_p^{GT})$.

A Gaussian deformation will have almost zero motion vector contribution at the image boundary due to Z-motion and hence diffusion of reliable flow from the boundary will not take place to the image center. On the contrary, a 2D



(a) (b)

Fig. 3: (a) Reconstruction of a Gaussian deformation($error_{3D}$=34%),(b) Reconstruction of a 2D parabola($error_{3D}$=3.98%) (c) 3D reconstruction error plot for varying noise levels. (Red- Ground truth deformation, blue-recovered deformation)

parabola will have maximum contribution to 2D motion on the boundaries and hence Z-motion should be recovered with high accuracy. This is evident from figure 3(a) and (b) where the Gaussian deformation has an error of 34.04% and the parabolic deformation has only error of 3.98%.

5.3 Real images

We tested our algorithm with the paper bending sequence and paper creasing sequence from [11]. We used the Multi-Frame Subspace Flow [5], [6] for optical flow computation between frames. For real images where initial rigid configuration is not available, we assume the object to be planar and at an arbitrary distance of 320mm from the camera. The results for real experiments are shown in figure 4. Results show that our method can successfully recover the shape for smooth non-rigid deformation. This is possible because of the diffusion of Z-motion from the boundaries to the image center due to the presence of smoothness term in the local-global optimization. We weighted the Z-motion term with a radially distributed weighting function with maximum value at the image boundary and minimum value at the image center, which emphasizes that the reliability of the Z-motion is low at the image center and high at the boundary. Even for sharp deformations like the creasing paper, the proposed algorithm could recover 3D deformation that matches the target image.

6. Conclusion

In this paper we have shown that the reliability of estimation of 3D flow from 2D flow is fundamentally dependent on the spatial location of the image. We have illustrated our hypothesis for Gaussian and Parabolic synthetic diffusion case. Although we have weighted the Z-motion term in real image deformation estimation with a radially outwards parabolic weighing function to produce realistic deformations on real images, this is an ad-hoc method and further study is required to exploit this spatial dependency information for improved 3D flow estimation.

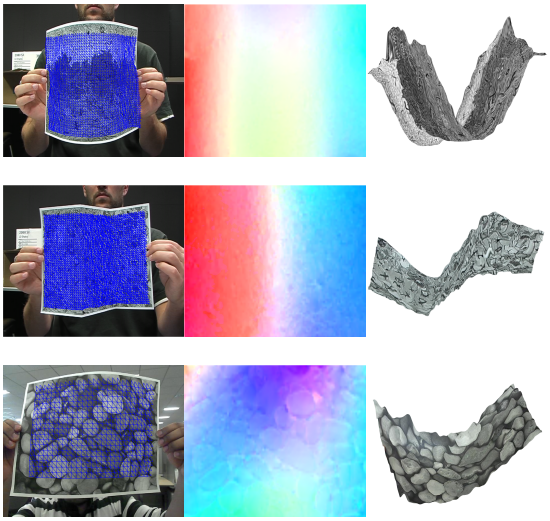


Fig. 4: Results from the real dataset. From left to right column: Image, Optical flow visualization, Texture mapped 3D deformation.

References

[1] Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J. and Szeliski, R.: A Database and Evaluation Methodology for Optical Flow, *International Journal of Computer Vision*, Vol. 92, No. 1, pp. 1–31 (2010).

[2] Bartoli, A., Grard, Y., Chadebecq, F., Collins, T. and Pizarro, D.: Shape-from-Template, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 10, pp. 2099–2118 (2015).

[3] Bruhn, A., Weickert, J. and Schnörr, C.: Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods, *International Journal of Computer Vision*, Vol. 61, No. 3, pp. 211–231.

[4] Garg, R., Roussos, A. and Agapito, L.: Dense Variational Reconstruction of Non-rigid Surfaces from Monocular Video, *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 1272–1279 (2013).

[5] Garg, R., Roussos, A. and Agapito, L.: Robust Trajectory-space TV-L1 Optical Flow for Non-rigid Sequences, *Proceedings of the 8th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, EMMCVPR'11, Berlin, Heidelberg, Springer-Verlag, pp. 300–314 (2011).

[6] Garg, R., Roussos, A. and Agapito, L.: A Variational Approach to Video Registration with Subspace Constraints, *International Journal of Computer Vision*, Vol. 104, No. 3, pp. 286–314 (2013).

[7] Horn, B. K. and Schunck, B. G.: Determining Optical Flow, Technical report, Cambridge, MA, USA (1980).

[8] Liu, C., Freeman, W. T. and Adelson, E. H.: Beyond pixels: exploring new representations and applications for motion analysis, PhD Thesis, Massachusetts Institute of Technology (2009).

[9] Lucas, B. D. and Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision, *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., pp. 674–679 (1981).

[10] Lucas, B. D.: Generalized Image Matching by the Method of Differences, PhD Thesis, Pittsburgh, PA, USA (1985). AAI8601180.

[11] Salzmann, M., Hartley, R. and Fua, P.: Convex optimization for deformable surface 3-d tracking, *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, IEEE, pp. 1–8 (2007).

[12] Salzmann, M., Moreno-Noguer, F., Lepetit, V. and Fua, P.: Closed-form solution to non-rigid 3D surface registration, *Computer Vision–ECCV 2008*, Springer, pp. 581–594 (2008).

[13] Salzmann, M., Urtasun, R. and Fua, P.: Local deformation models for monocular 3D shape recovery, *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, IEEE, pp. 1–8 (2008).