A Tensor Voting Approach for Multi-View 3D Scene Flow Estimation and Refinement

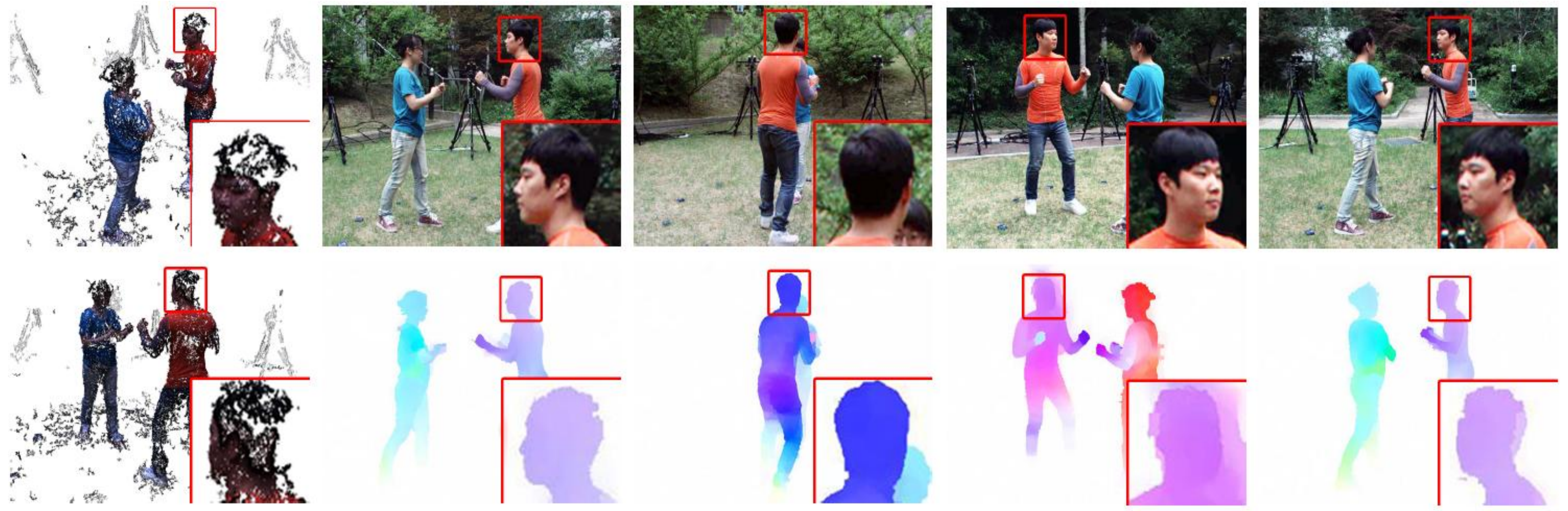
Jaesik Park Tae Hyun Oh Jiyoung Jung Yu-Wing Tai In So Kweon

Korea Advanced Institute of Science and Technology (KAIST)

Introduction

- **Scene flow** is a dense 3D motion vector field which describes motion of objects in 3D world.
- There are many previous works use stereoscopic inputs whereas the estimated scene flow is 2.5D; In this work, we use multi-view data to estimate a complete 3D scene flow.
- We extends the **closed form tensor voting** (Wu *et al*. CVPR2010) to scene flow estimation and refinement
- Our approach processes directly on 3D point cloud which is model free and memory efficient

Scene Flow Estimation



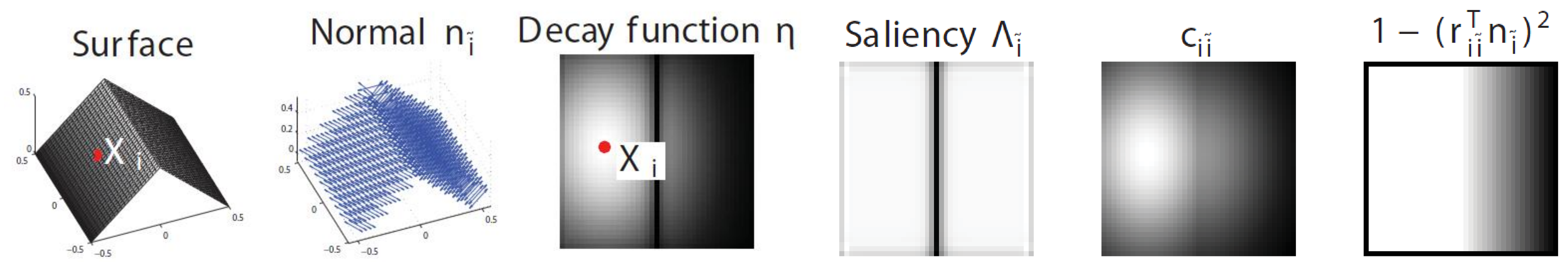
PMVS First row : Input image, Second row : 2D Optical flow

We estimate 3D geometry $\mathbf{X}_i$ and 3D scene flow $\mathbf{F}_i$ by using

- Patch based multi-view stereo(PMVS) (Furukawa *et al*. PAMI 2010)
- 2D Optical flow (Sun *et al*. CVPR 2010)
- Triangulation using direct linear transform (Vedula *et al*. PAMI 2005)

This method gives reasonable initialization, but not accurate enough.

Refinement of estimated Scene Flow



We regard $\mathbf{F}_i = m_i \vec{\mathbf{F}_i}$ and refine it by two steps:

Step 1: Refine direction vector, $\vec{\mathbf{F}_i}$ using closed form tensor voting[19,25]

$$\mathbf{S}_{i\tilde{i}} = \eta(\mathbf{X}_i, \mathbf{X}_{\tilde{i}}, \mathbf{n}_i, \mathbf{n}_{\tilde{i}}) \vec{\mathbf{F}_{\tilde{i}}} \vec{\mathbf{F}_{\tilde{i}}}^T$$

$$\eta(\mathbf{X}_i, \mathbf{X}_{\tilde{i}}, \mathbf{n}_i, \mathbf{n}_{\tilde{i}}) = c_{i\tilde{i}}[\Lambda_i(1-(\mathbf{r}_{i\tilde{i}}^T \mathbf{n}_i)^2) + \Lambda_{\tilde{i}}(1-(\mathbf{r}_{i\tilde{i}}^T \mathbf{n}_{\tilde{i}})^2)]$$

where $c_{i\tilde{i}}$ is Euclidean distance and $\Lambda_i = \lambda_{1,i} - \lambda_{2,i}$ is surface saliency of $\mathbf{X}_i$. $\Lambda_{\tilde{i}}$ is also defined in similar manner.

Step 2: Refine magnitude m_i by minimizing

$$E(m_i) = \frac{1}{\mathbf{X}_{\sigma,i}^2} \left\| \mathbf{X}_{\mu,i} - (\mathbf{X}_i + m_i \vec{\mathbf{F}_i}) \right\|^2 + \kappa \sum_{\tilde{i}} \eta(\mathbf{X}_i, \mathbf{X}_{\tilde{i}}, \mathbf{n}_i, \mathbf{n}_{\tilde{i}}) \left\| m_i - m_{\tilde{i}} \right\|^2$$

where $\mathbf{X}_{\mu,i}$ and $\mathbf{X}_{\sigma,i}^2$ are mean and variance of 3D points which are adjacent to the ray $(\mathbf{X}_i + m_i \vec{\mathbf{F}_i})$.

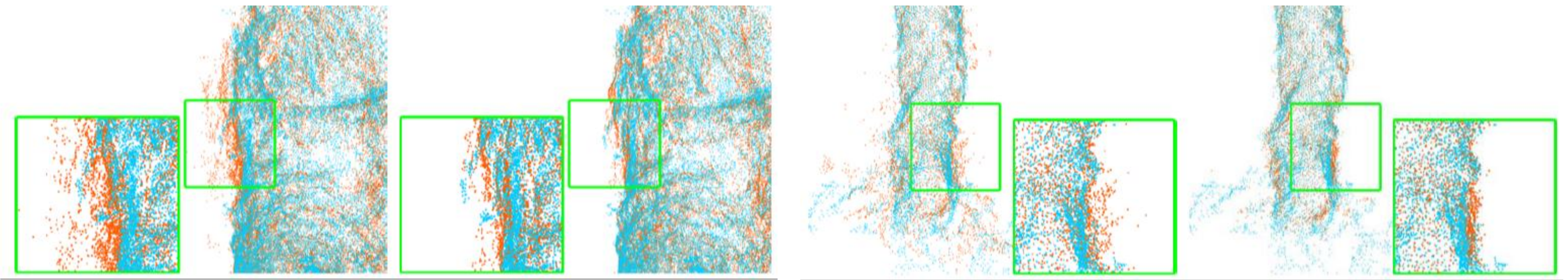
Note: to improve temporal coherency of the estimated scene flow, the voting field includes tensors in t-1, t+1 frames.

References

Vedula, S., Baker, S., Rander, P., Collins, R., Kanade, T.: Three-dimensional scene flow. IEEE Trans. on PAMI, 2005
Basha, T., Aviv, T., Moses, Y., Kiryati, N.: Multi-view scene flow estimation : A view centered variational approach. In: CVPR. (2010)
Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. In: ICCV. (2007)
Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans. on PAMI 32(8) (2010) 1362-1376
Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: CVPR. (2010)
Wu, T.P., Yeung, S.K., Jia, J., Tang, C.K.: Quasi-dense 3d reconstruction using tensor-based multiview stereo. In: CVPR. (2010)

Experiment on Real-world Dataset

Validation #1 : Point Clouds Overlapping using Refined flow



Baseline Ours Baseline Ours

Validation #2 : 3D Reconstruction using Refined Flow

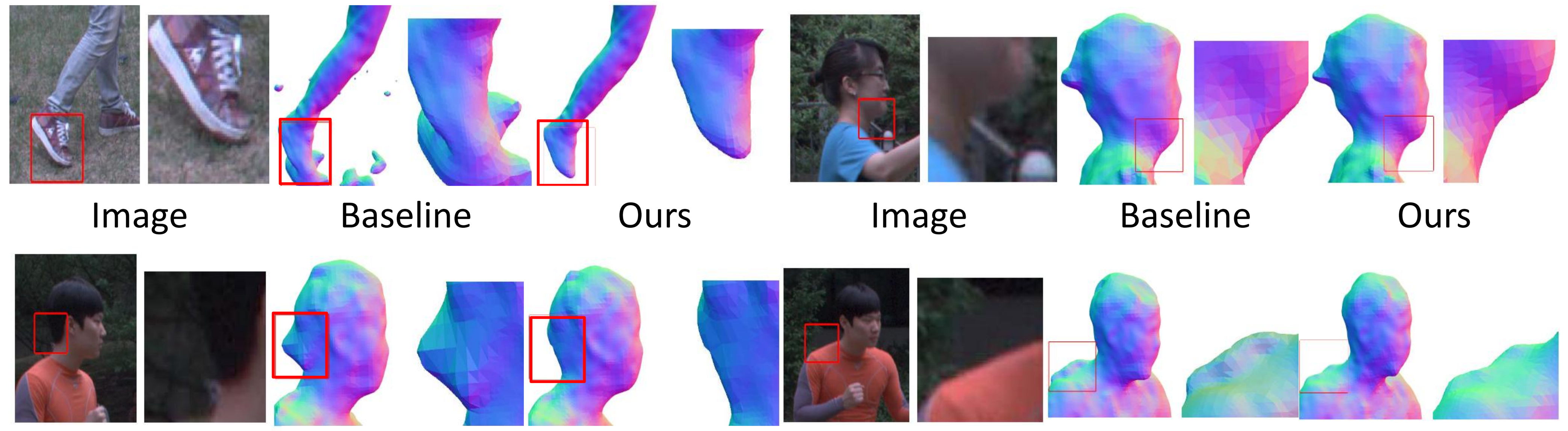
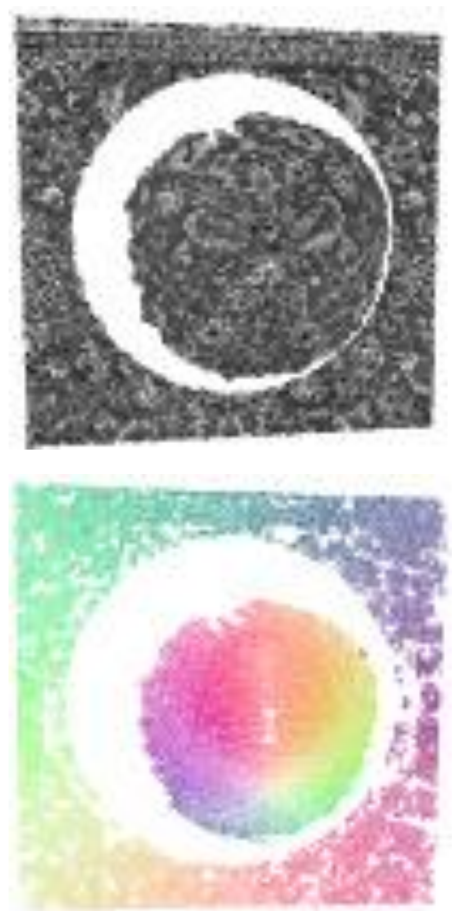


Image Baseline Ours Image Baseline Ours

Image Baseline Ours Image Baseline Ours

Experiment on Synthetic Dataset



Method	Measurement	$NRMS_{\mathbf{X}}$(%)	$NRMS_{\mathbf{F}}$(%)	$AAE_{\mathbf{F}}$(deg)
Huguet *et al.* (ICCV2007)	w/o Discontinuities w/o Occlusions All pixels	9.82 1.19 10.43	15.96 11.04 19.09	7.17 6.66 9.20
Basha *et al.* (CVPR2010)	w/o Discontinuities w/o Occlusions All pixels	0.65 1.99 4.39	**2.94** 5.63 9.71	**1.32** **2.09** 3.39
Ours (Baseline)	w/o Discontinuities w/o Occlusions All pixels	0.25 0.26 1.12	6.43 6.99 7.89	4.74 4.98 5.28
Ours (After Refinement)	w/o Discontinuities w/o Occlusions All pixels	**0.23** **0.24** **0.57**	4.88 **5.07** **5.42**	2.73 2.72 **2.83**

NRMS* : normalized root mean square, **AAE : Absolute angular error

This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No.2012-0000986 and No. 2012-0003359) and partially by Microsoft Research Asia under KAIST-Microsoft Research Collaboration Center(KMCC). Real world dataset was provided by ETRI, Korea.