On the Evaluation of Scene Flow Estimation

Philippos Mordohai

Stevens Institute of Technology
mordohai@cs.stevens.edu

Abstract. This paper surveys the state of the art in evaluating the performance of scene flow estimation and points out the difficulties in generating benchmarks with ground truth which have not allowed the development of general, reliable solutions. Hopefully, the renewed interest in dynamic 3D content, which has led to increased research in this area, will also lead to more rigorous evaluation and more effective algorithms. We begin by classifying methods that estimate depth, motion or both from multi-view sequences according to their parameterization of shape and motion. Then, we present several criteria for their evaluation, discuss their strengths and weaknesses and conclude with recommendations.

1 Introduction

Multiple-view reconstruction has been one of the most active areas in computer vision and, as a result, impressive 3D models can now be generated for individual objects [1–3] and large scale scenes [4–6]. Moreover, commercial entities, including Apple, Google, Nokia and Acute3D, have been able to produce accurate models from massive amounts of images and video, proving that general solutions in uncontrolled environments are possible.

This, however, is not the case for the reconstruction of dynamic scenes despite the broader spectrum of applications that would be made possible. In contrast to static 3D models used for visualization, 3D mapping, virtual tourism and measuring distances, 3D reconstructions of dynamic scenes can reach significantly more people. Applications include free-viewpoint video for 3D TV, films or video-games; markerless motion capture for entertainment, clinical medicine and biomechanical analysis; and dynamic augmented reality.

The problem of estimating 3D shape and motion from images was named *scene flow estimation* by Vedula et al. [7] who presented the first analysis of what can be estimated depending on how much about the scene is known. After a few early publications [7–9] interest in this topic waned, until recently, when it was sparked again by the growing demand for 3D content. Significant progress has been made but the problem has not been fully solved. Much better results have been obtained for humans via the use of articulated models which drastically reduces the number of parameters to be estimated [10–13]. Due to the very different nature of such strongly model-based estimation, we consider these methods out of scope here. We will, however, consider methods that do not explicitly compute motion, such as space-time stereo [14, 15], as long as images from different time instances are used in the computation of depth for a given frame[1].

In Section 2, we present a taxonomy of the most relevant methods according to their parameterization of shape and motion. Then, in Section 3, we examine a number of criteria that have been used for evaluating these methods and discuss their strengths and weaknesses. Unlike other problems in computer vision, generating datasets with ground truth for scene flow estimation is not straightforward due to the unavailability of a suitable technology. Overcoming this challenge and generating widely used benchmarks will be instrumental for progress in this area, as it was for related problems, such as binocular [16] and multi-view [17, 18] stereo and optical flow [19].

2 A Taxonomy of the Methods

In this section, we classify dense or quasi-dense scene flow methods according to their parameterization. For example, viewpoint-based methods may parameterize each pixel using four parameters: one for depth on the ray and three for 3D motion to the next frame. On the other hand, a world-based method may endow a 3D point with six parameters: three for translational and three for rotational motion.

2.1 Temporally-supported Depth

We begin with methods that take into account frames captured at different times to estimate depth for the pixels of the current frame. These methods do not explicitly estimate motion, because this is either not needed for the application at hand, or the computational cost of evaluating photoconsistency in 4D state spaces is too high. Space-time stereo [14, 20, 15, 21–24] falls under this category. We note here that motion must be very small for the assumptions made by these methods to remain valid.

Methods that can handle larger motions [25–27] use optical flow to identify temporal correspondences across frames and compute the matching cost aggregating information over time. The limitations in both cases are the assumption that depth remains constant over time and that 3D motion is not estimated.

2.2 Temporally-supported 3D Shape

Dynamic shapes can be extracted as 3D iso-surfaces in 4D spatio-temporal volumes [28, 29] or as the boundary facets of 4D Delaunay meshes [30]. A weaker form of such temporally consistent shape estimation is based on minimizing the distance between surfaces at consecutive time frames [31]. These methods are limited to sequences with very slow motion so that the bounding surfaces of the 4D volumes are smooth, while the resulting temporal correspondences are set-wise and not point-wise.

[1] We use the term *frames* to refer to images taken at different times.

2.3 Depth and Motion per Pixel

This category includes viewpoint-based methods that estimate shape and 3D motion (four degrees of freedom) for each pixel of the reference view. Due to the size of the problem, early work was based on fitting parametric motion models to image segments thus drastically reducing the number of degrees of freedom to be estimated [32–34]. Isard and MacCormick [35] proposed an MRF with a 5D state space, including a flag for occlusion. Variational approaches [36–40] initialize shape by stereo matching and then estimate shape and motion jointly until convergence to the nearest local minimum of the objective function. Other methods explore the 4D search space using a winner-take-all scheme or dynamic programming [41] or by growing correspondence seeds [42].

To reduce computational complexity, several authors decouple depth and motion estimation [43–46]. Depth estimation, however, is still affected by temporal correspondences as in the methods of Section 2.1 or by predictions made using scene flow estimates of the previous frame.

2.4 3D Shape and Motion

This category includes methods that employ world-based representations and thus are not limited to $2\frac{1}{2}$-D surfaces. Initially, Vedula et al. [7] formulated and analyzed three algorithms depending on the degree to which scene geometry is known. In a separate paper, Vedula et al. [47] extended space carving to the 6D space of all shapes and flows. Other representations for joint shape and motion estimation include surface patches [8], subdivision surfaces [9], a hybrid of oriented 3D points and signed distance functions [48], probabilistic occupancy-motion grids [49] and a collection of rigid parts discovered via EM [50].

Furukawa and Ponce [51, 52] and Courchay et al. [53] estimate the scene flow of a single mesh with fixed topology. Six degrees of freedom are estimated per vertex to capture both translation and rotation.

Shape and motion estimation can be loosely coupled by reconstructing an initial 3D shape, estimating motion to the next frame and using the motion estimates to predict the shape at the next frame. Pons et al. [3] apply a variational method that minimizes the prediction error of the shape and motion estimates. Popham et al. [54] track surface patches in 3D in long sequences. Li et al. [55] extract a watertight mesh from point clouds reconstructed by variational stereo and then address scene flow as volumetric deformation.

2.5 Motion of 3D Shape

The final class of methods estimate motion between consecutive observations of shape. In other words, these algorithms consider shape estimates as input and focus on establishing temporal correspondences. More common are methods for tracking meshes with fixed topology, which are either initialized using multi-view stereo on the first frame or are externally provided. Most of these approaches use reliable, sparse features of various types [31, 56–59] to guide dense correspondence for all points. Cagniart et al. [60] divide the surface into elementary patches which provide integration domains for increased tolerance to noise.

These methods estimate shape independently at each time frame without the benefit of temporal correspondences, which can improve accuracy (Section 3.1). Moreover, most of them, except [60], are restricted to a single watertight mesh.

3 Evaluation Techniques

We examine several criteria that have been used to gauge the performance of scene flow estimation in the literature.

3.1 Ground Truth Depth

The easiest aspect of scene flow estimation to be evaluated is the accuracy of depth estimation. What should be measured is the improvement due to the inclusion of temporal correspondences and not the performance of a single-frame version of the algorithm on static benchmarks. Sizintsev and Wildes [24, 46] are the only authors, to our knowledge, to perform such experiments by having cameras and scenes on motorized stages and by acquiring ground truth using structured light. These experiments show that the use of temporal information improves reconstruction accuracy. Other authors have performed similar experiments on synthetic data [26, 22, 23], also showing improvement.

Generating ground truth for this type of tests is feasible using LIDAR or consumer depth cameras. The latter have the advantage of capturing depth for all pixels in a single shot, while LIDAR has larger range but scans points sequentially. (Flash LIDAR does not appear to be a viable solution yet.) Despite the sparsity of LIDAR measurements, their use is fair since sampling is unbiased and no markers which would aid the algorithm under evaluation are used.

3.2 Ground Truth 2D Motion

Benchmarking motion estimation, even in 2D, is considerably harder than stereo. Baker et al. [61] had to paint hidden fluorescent texture on the scene to generate ground truth optical flow for the Middlebury Optical Flow evaluation. This texture is only visible to special sensors that were used along with regular cameras. Only Li et al. [55] have measured 2D motion errors using real inputs with ground truth by clicking a small number of features in a long sequence. Quantitative optical flow results on synthetic data have been presented on a sphere [42] and a humanoid model captured by 16 virtual cameras [57].

Generating data with ground truth motion is not scalable since it requires considerable manual effort in the form of painting or clicking. The latter suffers from significant selection bias since only distinctive points can be reliably localized by the annotators, but these are also points on which algorithms perform well. Further, it appears that with some additional effort one could generate

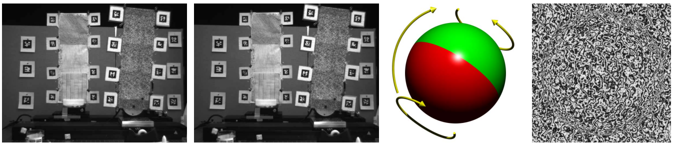


Fig. 1. Left: two frames from the left camera of the ground truth data captured by Sizintsev and Wildes [46]. The surfaces move independently, but the markers simplify matching. Right: The synthetic sphere of Huguet and Devernay [36]: a diagram showing that the two hemispheres undergo separate rotations and the texture map.

ground truth for 3D shape and motion by extending the annotation to more than one camera or by adding a second special sensor.

3.3 Ground Truth Scene Flow

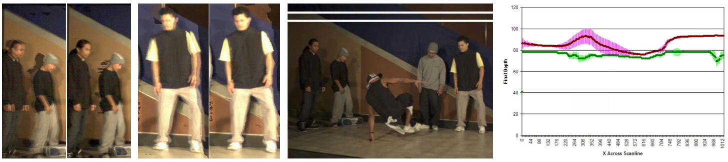
Sizintsev and Wildes [46] present a quantitative evaluation of dense scene flow using the motorized stage described above to generate ground truth. While it is unprecedented, this experiment is not ideal since several fiducial markers had to be placed on each independently moving surface to aid motion estimation (Fig. 1). This also improves the accuracy of the algorithms being evaluated, not only on the markers themselves, but also on other pixels. Popham et al. [54] evaluated the accuracy of their algorithm over long sequences (90 frames) on a small number of points manually clicked on two images for each time frame. While the reported accuracy is high, this approach does not scale well and also suffers from selection bias as that of Li et al. [55].

The majority of scene flow estimation algorithms that include quantitative results obtain them on synthetic datasets, with the most popular being the one by Huguet and Devernay [36] (Fig. 1) which has been used by several authors [36, 38, 40, 45]. Other options include synthetic spheres, ellipses or cylinders [41, 49, 39, 42, 50, 45] or synthetic scenes [43, 45].

3.4 Image Prediction

According to Szeliski [62] and Kilner et al. [63], one can leave out one of the cameras and try to predict its images using all other available data. This test may not reveal errors in areas of low texture, but is ideal for evaluating novel view synthesis. Taking this approach a step further, Kilner et al. [63] showed that there is significant correlation between the prediction error on an existing view and the differences obtained by rendering two different real images to a virtual camera using the reconstructed surface to determine the mapping. Therefore, this test can be performed without "sacrificing" one of the input cameras, thus degrading the quality of the estimation.

To the best of our knowledge, our previous work [25] is one of the few examples in which this technique was applied. Using the calibrated and synchronized



(a) Image prediction (b) Background conistency

Fig. 2. Evaluation of the approach of Larsen et al. [25] on the dataset of Zitnick et al. [64]. (a) Image prediction by rendering colored depth map on a different view. Zoomed in results of [64] and [25] are shown on the left and right respectively. (b) A frame of the breakdancing sequence with row 40 highlighted and a plot of the median depth and standard deviation for all pixels under the approach of [64] in purple (top curve) and the approach of [25] in green (bottom curve). The latter has significantly smaller standard deviation for the depth of each pixel.

videos provided by Microsoft Research as input and the associated reconstructions [64] as the baseline, we were able to show that temporal consistency leads to lower reprojection errors. See Fig. 2(a) for some results.

Here, we propose to extend this criterion to consider not only prediction of different viewpoints, but also of images taken at the following time step using the estimated scene flow. This type of evaluation requires minimal effort compared to the alternatives and is applicable to most methods, in particular mesh tracking methods, for which evaluation of shape is pointless. Despite not capturing errors in textureless regions well, this criterion is well-suited for applications that value novel view synthesis more than metric accuracy.

3.5 Temporal Consistency of the Background

A different criterion [25] is to evaluate the temporal consistency of at least the static parts of the scene by selecting pixels that remain static throughout the sequence and measuring the variance of the estimated depth. Small variance indicates that the algorithm maintains a consistent depth, regardless of its precision, and thus generates results with reduced jittering. We argued that for many applications, a constant, but slightly wrong, depth for the background is more visually pleasing than more accurate depth that fluctuates over time. Some results for the breakdancing sequence of [64] are shown in Fig. 2(b).

The limitation of this method is that it is inapplicable to the foreground, as well as to pixels of the background whose visibility changes during the sequence. It is also inapplicable to methods that require silhouettes as inputs.

3.6 Forward-Backward Consistency

In the absence of ground truth data, Furukawa and Ponce [51] concatenated forward and reverse videos around a common frame, creating sequences such as $f_1 f_2 f_3 f_2 f_1$, and then measured the consistency of scene flow estimates between the same pairs of frames that appear in reverse order, such as $f_1 f_2$ and $f_2 f_1$.

Ideally, shape estimates should be identical and motion vectors should have the same magnitude but opposite orientation. This technique requires further analysis since certain estimators may produce consistent, but erroneous, results. On the other hand, the ease of creating the input data is a strength.

3.7 Scenes with Ground Truth Depth undergoing Rigid Motion

The last technique of this section is to generate test sequences by acquiring videos with a moving camera rig of static scenes with ground truth depth. The ground truth depth can be acquired separately using LIDAR or consumer depth cameras and motion between frames can be estimated using the Iterative Closest Point algorithm or structure from motion. The output of this procedure is not only depth for all frames, but also ground truth dense scene flow since exact rigid transformations can be computed for all points that are visible in the range scans. As long as the algorithm under evaluation is not aware of the global rigidity of the motion and thus does not enforce the relevant constraints, the evaluation is both fair and informative. A simple form of this criterion was used by Huguet and Devernay [36] and Liu and Philomin [37] on the Middlebury data [16] using four images of each scene. Scene flow was evaluated between binocular pairs without taking into account that the true motion is pure horizontal translation. Prediction errors were also included in [37].

This technique appears to provide a good trade-off between effort required to generate the data and thoroughness of the evaluation, since, unlike image prediction, the accuracy over al pixels, textured or not, can be measured. There are at least two limitations: the technique is better suited to binocular or other narrow-baseline methods due to potential difficulties in moving a large camera rig rigidly; and the data may be unsuitable for methods that require silhouettes.

4 Conclusions

After surveying the state of the art in scene flow estimation and evaluation, the main conclusion is that the problem remains unsolved, but that it is not “unsolvable”. In fact, great progress has been made in the last 3-5 years on the algorithmic front, but evaluation is still lacking.

While the generation of synthetic video sequences using proper rendering techniques seems the most practical way of producing data with ground truth, results on such sequences should be interpreted with caution. It is often very hard to capture all failure modes of stereo matching and motion estimation on synthetic data for a number of reasons including issues related to noise modeling or irregularities of camera response functions. After evaluating the criteria of Section 3, we propose the use of a combination of *image prediction* errors, which require little effort to measure and are well-suited for view synthesis tasks, and *scenes with ground truth depth undergoing rigid motions*, which require non-trivial effort to generate but capture all aspects of scene flow estimation. The *forward-backward consistency* technique is also worth further investigation.

Acknowledgement: This work was supported in part by NSF award 1217797.

References

1. Hernández Esteban, C., Schmitt, F.: Silhouette and stereo fusion for 3D object modeling. CVIU **96** (2004) 367–392
2. Furukawa, Y., Ponce, J.: Carved visual hulls for imagebased modeling. In: ECCV. (2006) I: 564–577
3. Pons, J.P., Keriven, R., Faugeras, O.D.: Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. IJCV **72** (2007) 179–193
4. Furukawa, Y., Curless, B., Seitz, S., Szeliski, R.: Towards internet-scale multi-view stereo. In: CVPR. (2010)
5. Frahm, J., Fite Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y., Dunn, E., Clipp, B., Lazebnik, S., Pollefeys, M.: Building rome on a cloudless day. In: ECCV. (2010) IV: 368–381
6. Vu, H., Labatut, P., Pons, J.P., Keriven, R.: High accuracy and visibility-consistent dense multi-view stereo. PAMI (2011)
7. Vedula, S., Baker, S., Rander, P., Collins, R.T., Kanade, T.: Three-dimensional scene flow. In: ICCV. (1999) 722–729
8. Carceroni, R.L., Kutulakos, K.N.: Multi-view scene capture by surfel sampling: From video streams to non-rigid 3D motion, shape and reflectance. IJCV **49** (2002) 175–214
9. Neumann, J., Aloimonos, Y.: Spatio-temporal stereo using multi-resolution subdivision surfaces. IJCV **47** (2002) 181–193
10. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: Shape completion and animation of people. ACM Trans. on Graphics **24** (2005) 408–416
11. Cheung, K.M., Baker, S., Kanade, T.: Shape-from-silhouette across time part ii: Applications to human modeling and markerless motion tracking. IJCV **63** (2005) 225–245
12. Starck, J., Hilton, A.: Surface capture for performance-based animation. IEEE Computer Graphics and Applications **27** (2007) 21–31
13. Vlasic, D., Baran, I., Matusik, W., Popović, J.: Articulated mesh animation from multi-view silhouettes. ACM Trans. on Graphics **27** (2008) 1–9
14. Zhang, L., Curless, B., Seitz, S.M.: Spacetime stereo: shape recovery for dynamic scenes. In: CVPR. (2003) II: 367–374
15. Davis, J., Nehab, D., Ramamoorthi, R., Rusinkiewicz, S.: Spacetime stereo: A unifying framework for depth from triangulation. PAMI **27** (2005) 296–302
16. Scharstein, D., Szeliski, R.S.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV **47** (2002) 7–42
17. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: CVPR. (2006) 519–528
18. Strecha, C., von Hansen, W., Van Gool, L.J., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: CVPR. (2008)
19. Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., Szeliski, R.: A database and evaluation methodology for optical flow. IJCV **92** (2011) 1–31
20. Leung, C., Appleton, B., Lovell, B.C., Sun, C.: An energy minimisation approach to stereo-temporal dense reconstruction. In: ICPR. (2004) IV: 72–75

21. Williams, O., Isard, M., MacCormick, J.: Estimating disparity and occlusions in stereo video sequences. In: CVPR. (2005) II:250–257
22. Richardt, C., Orr, D., Davies, I., Criminisi, A., Dodgson, N.: Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid. In: ECCV. (2010) III: 510–523
23. Hosni, A., Rhemann, C., Bleyer, M., Gelautz, M.: Temporally consistent disparity and optical flow via efficient spatio-temporal filtering. In: PSIVT. (2011) I: 165–177
24. Sizintsev, M., Wildes, R.: Spatiotemporal oriented energies for spacetime stereo. In: ICCV. (2011) 1140–1147
25. Larsen, E.S., Mordohai, P., Pollefeys, M., Fuchs, H.: Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. In: ICCV. (2007)
26. Bartczak, B., Jung, D., Koch, R.: Real-time neighborhood based disparity estimation incorporating temporal evidence. In: DAGM. (2008) 153–162
27. Yang, W., Zhang, G., Bao, H., Kim, J., Lee, H.Y.: Consistent depth maps recovery from a trinocular video sequence. In: CVPR. (2012)
28. Goldluecke, B., Magnor, M.: Space-time isosurface evolution for temporally coherent 3D reconstruction. In: CVPR. (2004) 350–355
29. Sharf, A., Alcantara, D.A., Lewiner, T., Greif, C., Sheffer, A., Amenta, N., Cohen-Or, D.: Space-time surface reconstruction using incompressible flow. ACM Trans. on Graphics **27** (2008) 1–10
30. Aganj, E., Pons, J.P., Segonne, F., Keriven, R.: Spatio-temporal shape from silhouette using four-dimensional delaunay meshing. In: ICCV. (2007)
31. Starck, J., Hilton, A.: Correspondence labelling for wide-timeframe free-form surface matching. In: ICCV. (2007)
32. Tao, H., Sawhney, H.S., Kumar, R.: Dynamic depth recovery from multiple synchronized video streams. In: CVPR. (2001) I:118–124
33. Zhang, Y., Kambhamettu, C.: On 3-d scene flow and structure recovery from multiview image sequences. PAMI **33** (2003) 592–606
34. Li, R., Sclaroff, S.: Multi-scale 3D scene flow from binocular stereo sequences. CVIU **110** (2008) 75–90
35. Isard, M., MacCormick, J.P.: Dense motion and disparity estimation via loopy belief propagation. In: ACCV. (2006) II:32–41
36. Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. In: ICCV. (2007)
37. Liu, F., Philomin, V.: Disparity estimation in stereo sequences using scene flow. In: BMVC. (2009)
38. Basha, T., Moses, Y., Kiryati, N.: Multi-view scene flow estimation: A view centered variational approach. In: CVPR. (2010)
39. Valgaerts, L., Bruhn, A., Zimmer, H., Weickert, J., Stoll, C., Theobalt, C.: Joint estimation of motion, structure and geometry from stereo sequences. In: ECCV. (2010)
40. Vogel, C., Schindler, K., Roth, S.: 3d scene flow estimation with a rigid motion prior. In: ICCV. (2011) 1291–1298
41. Gong, M.: Real-time joint disparity and disparity flow estimation on programmable graphics hardware. CVIU **113** (2009) 90 – 100
42. Cech, J., Sanchez-Riera, J., Horaud, R.: Scene flow estimation by growing correspondence seeds. In: CVPR. (2011)
43. Rabe, C., Müller, T., Wedel, A., Franke, U.: Dense, robust, and accurate motion field estimation from stereo image sequences in real-time. In: ECCV. (2010) IV: 582–595

44. Müller, T., Rannacher, J., Rabe, C., Franke, U.: Feature- and depth-supported modified total variation optical flow for 3d motion field estimation in real scenes. In: CVPR. (2011)
45. Wedel, A., Brox, T., Vaudrey, T., Rabe, C., Franke, U., Cremers, D.: Stereoscopic scene flow computation for 3d motion understanding. IJCV **95** (2011) 29–51
46. Sizintsev, M., Wildes, R.: Spatiotemporal stereo and scene flow via stequel matching. PAMI **34** (2012) 1206–1219
47. Vedula, S., Baker, S., Seitz, S.M., Kanade, T.: Shape and motion carving in 6D. In: CVPR. (2000) 592–598
48. Kwatra, V., Mordohai, P., Kumar Penta, S., Narain, R., Carlson, M., Pollefeys, M., Lin, M.: Fluid in video: Augmenting real video with simulated fluids. Computer Graphics Forum **27** (2008) 487–496
49. Guan, L., Franco, J.S., Boyer, E., Pollefeys, M.: Probabilistic 3D occupancy flow with latent silhouette cues. In: CVPR. (2010)
50. Franco, J.S., Boyer, E.: Learning temporally consistent rigidities. In: CVPR. (2011)
51. Furukawa, Y., Ponce, J.: Dense 3D motion capture from synchronized video streams. In: CVPR. (2008)
52. Furukawa, Y., Ponce, J.: Dense 3D motion capture for human faces. In: CVPR. (2009)
53. Courchay, J., Pons, J.P., Monasse, P., Keriven, R.: Dense and accurate spatio-temporal multi-view stereovision. In: ACCV. (2009) II: 11–22
54. Popham, T., Bhalerao, A., Wilson, R.: Multi-frame scene-flow estimation using a patch model and smooth motion prior. In: BMVC Wkhp. (2010)
55. Li, K., Dai, Q., Xu, W.: Markerless shape and motion capture from multiview video sequences. IEEE Trans. on Circuits and Systems for Video Technology **21** (2011) 320–334
56. Ahmed, N., Theobalt, C., Rossl, C., Thrun, S., Seidel, H.P.: Dense correspondence finding for parametrization-free animation reconstruction from video. In: CVPR. (2008)
57. Varanasi, K., Zaharescu, A., Boyer, E., Horaud, R.: Temporal surface tracking using mesh evolution. In: ECCV. (2008) II: 30–43
58. Zeng, Y., Wang, C., Wang, Y., Gu, X., Samaras, D., Paragios, N.: Dense non-rigid surface registration using high-order graph matching. In: CVPR. (2010) 382–389
59. Huang, P., Hilton, A., Budd, C.: Global temporal registration of multiple non-rigid surface sequences. In: CVPR. (2011)
60. Cagniart, C., Boyer, E., Ilic, S.: Free-form mesh tracking: a patch-based approach. In: CVPR. (2010)
61. Baker, S., Roth, S., Scharstein, D., Black, M.J., Lewis, J.P., Szeliski, R.: A database and evaluation methodology for optical flow. In: ICCV. (2007)
62. Szeliski, R.: Prediction error as a quality metric for motion and stereo. In: ICCV. (1999) 781–788
63. Kilner, J., Starck, J., Guillemaut, J.Y., Hilton, A.: Objective quality assessment in free-viewpoint video production. Signal Processing: Image Communication **24** (2009) 3 – 16
64. Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.S.: High-quality video view interpolation using a layered representation. ACM Trans. on Graphics **23** (2004) 600–608