A Multimodal Approach for Image De-fencing and Depth Inpainting

Sankaraganesh Jonna*, Vikram S. Voleti*, Rajiv R. Sahay*, Mohan S. Kankanhalli†
*School of Information Technology and Department of Electrical Engineering, IIT Kharagpur, India
†School of Computing, National University of Singapore
{*sankar9.iitkgp, vikky2904, sahayiitm*}*@gmail.com, mohan@comp.nus.edu.sg*

***Abstract*—Low cost RGB-D sensors such as the Microsoft Kinect have enabled the use of depth data along with color images. In this work, we propose a multi-modal approach to address the problem of removal of fences/occlusions from images captured using a Kinect camera. We also perform depth completion by fusing data from multiple recorded depth maps affected by occlusions. The availability of aligned image and depth data from Kinect aids us in the detection of the fence locations. However, accurate estimation of the relative shifts between the captured color frames is necessary. Initially, for the case of static scene elements with simple relative motion between the camera and the objects, we propose the use of affine scale-invariant feature transform descriptor (ASIFT) to compute the relative global displacements. We also address the scenario wherein the relative motion between the frames may not be global using the depth map obtained by Kinect. For such a scenario involving complex motion of scene pixels, we use a recently proposed robust optical flow technique. We show results for challenging real-world data wherein the scene is dynamic. The inverse ill-posed problems of estimation of the de-fenced image and the inpainted depth map are solved using an optimization-based framework. Specifically, we model the unoccluded image and the completed depth map as two distinct Markov random fields, respectively, and obtain their maximum a-posteriori estimates using loopy belief propagation.**

***Keywords*—Image de-fencing, Inpainting, RGB-D data, Kinect, Belief propagation, Markov random field.**

I. Introduction

In recent years there has been a proliferation of smartphones/tablets which enable users to capture their cherished moments at any convenient time and location. Particularly, visitors to tourist destinations often feel hindered in capturing photographs/videos of objects that are occluded by barricades/fences for security purposes. Looking through grilled windows one often captures images of people, paintings or fragile antiquities. We show such a fenced scene in Fig. 1 (a). We address the problem of removal of such occlusions from these images/videos by taking a multi-modal approach in this work. We assume that the user has access to depth data along with color images. Such a constraint can easily be fulfilled today with the advent of sensors such as the Microsoft Kinect. For example, for the scene in Fig. 1 (a), we show the corresponding depth map captured by the Kinect sensor in Fig. 1 (b). Note that the fence depth profile is occluding the depth map of the background.

Time-of-flight sensors have been available since quite some time but they are not sufficiently accurate. Kinect uses an infrared based active triangulation approach and possesses better accuracy. The main advantage is that it has the depth sensor co-located with the RGB camera in the same physical apparatus. Hence, it is possible to obtain aligned depth maps as well as color images of the same scene. In this work, we leverage this fact to show that the depth map from Kinect can be used to

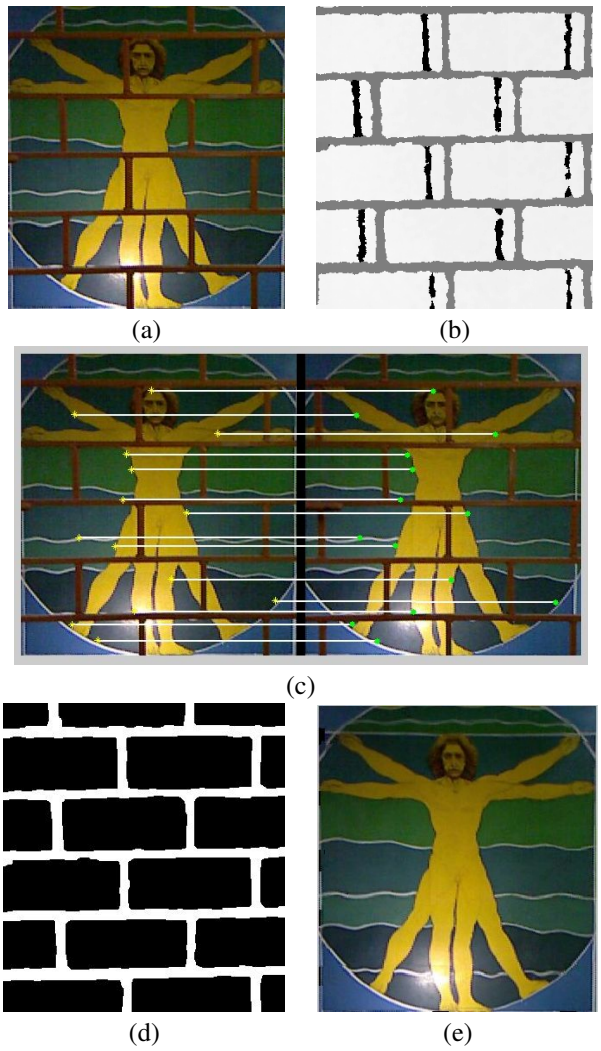


Fig. 1. (a) First observation obtained from the video. (b) Depth map corresponding to the first observation captured by the Kinect sensor. (c) Using ASIFT [13] descriptor to find corresponding points and hence estimating global dispacements between the first and second observations. (d) Fence mask corresponding to the first observation. (e) De-fenced image obtained with the proposed algorithm.

978-1-4799-7458-0/15/$31.00 ©2015 IEEE

detect occlusions such as fences/barricades.

Our multi-modal approach is related to prior work on image de-fencing using only image data [1], [2], [3]. The problem of image de-fencing and depth completion can be divided into three challenging tasks. Firstly, given an image/video of the occluded scene, the spatial locations of occlusions have to be estimated accurately. This is a non-trivial problem and has been addressed in the past [1], [2], [3]. However, unlike our multi-modal method the works of [1], [2], [3] address this task only with image data and do not have access to depth maps of the occluded scene. The major novelty of our work is that we have captured aligned depth data along with a color video of the scene using the Kinect sensor. Therefore, the sub-problem of fence detection is simplified considerably by processing in the depth data domain. Our work is also related to recent works in the area of depth inpainting [4], [5], [6], [7] but unlike these techniques our work uses the Kinect sensor to capture both RGB video and depth data.

Secondly, relative displacements between the image frames have to be estimated since we capture a video of the scene by translating the Kinect. Initially, we assume that the images are related by global shifts since the distance between the sensor and the occluded scene is significant. When the scene consists of two or more planes with considerable depth difference between them, the assumption of a single global displacement of the entire scene becomes invalid. For this case, we consider different motions for each plane. The depth map obtained from Kinect sensor proves useful again since it can be leveraged to segment out the different planes in the scene. The scenario is even more challenging when objects in the scene are dynamic or when the objects cannot be assumed to be planar.

These shifts between the image frames are critical to the success of our algorithm since some portion of the scene which is occluded in one frame will possibly be rendered visible in the other images due to the translation of the Kinect. In fact, this is the guiding principle behind our selection of frames from the captured RGB video and depth profiles using the Kinect sensor. Hence, our work is different from single image inpainting techniques [8], [9], [10], [11] which rely primarily on flowing information from surrounding regions, respecting isophotes, into the missing data areas. Traditional optical flow algorithms [12] yield inaccurate estimates due to the presence of missing image data at the occlusions. We use a robust image descriptor, affine scale invariant feature transform (ASIFT) [13], to obtain the spatial coordinates of the corresponding points between the relatively shifted frames. Sample points matched using this descriptor are shown in Fig. 1 (c). For more difficult cases, involving complex relative motion between the scene elements and the camera, we have used a recently proposed technique [14] for estimation of dense optical flow.

Thirdly, the observations have to be fused together so as to estimate a fence-free image and depth map. We model the input data obtained by the Kinect sensor and relate the unknown de-fenced image and inpainted depth profile with the captured RGB observations and depth maps, respectively. Since the estimation of the unoccluded image and completed depth profile are ill-posed problems, we use an optimization-based framework to estimate both unknown quantities. We model the de-fenced image and the inpainted depth profile as distinct Markov random fields, respectively, and obtain their maximum a-posteriori (MAP) estimates by minimizing suitably formulated objective functions.

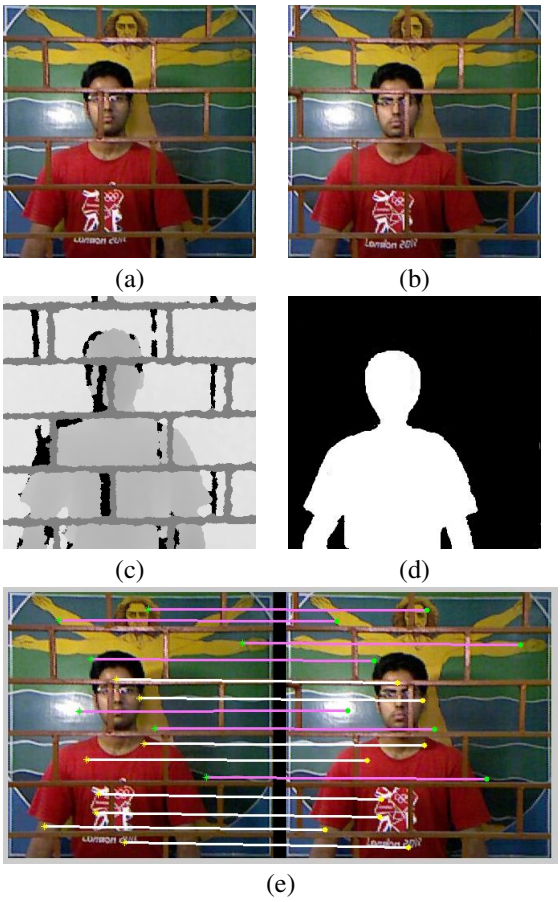


Fig. 2. (a), (b) First and second observations captured with the RGB camera. (c) Depth map corresponding to first observation obtained using Kinect. (d) Segmented foreground region using captured depth map. (e) Matching corresponding points in the foreground and background regions of the two observations using ASIFT [13] descriptor (White lines depict matches in the foreground and pink lines connect corresponding points in the background.)

Fig. 3. (a) De-fenced image obtained using our method for the case of non-global pixel motion. (b) Completed depth profile corresponding to the first observation.

II. Proposed Method

We propose to model the "de-fenced" image and the completed depth profile as two different Markov random fields. Initially, we seek to derive the maximum a-posteriori estimate of the de-fenced image given depth data and multiple frames from the captured video of the occluded scene using the

Microsoft Kinect sensor. The degradation model is given as

$$\mathbf{y}_m = O_m W_m \mathbf{x} + \mathbf{n}_m \tag{1}$$

where the operator O_m crops out the un-occluded pixels from the m^{th} frame, $\mathbf{y}_m$ represents the observations used, $\mathbf{x}$ is the de-fenced image, W_m is the warp matrix and $\mathbf{n}_m$ is the noise assumed as Gaussian.

We detect the fence pixels in every observation and therefore estimate O_m. Detection of fence pixels from image intensities only is a difficult problem addressed in the past in [15]. However, we show that it is possible to segment the depth map of the entire scene obtained using Kinect to localize the occlusions. We use Otsu's method [16] on the depth data to obtain the segmented fence masks. Initially, we assume that the non-fence pixels in the frames are shifted with respect to each other by a globally fixed amount. This assumption is justified if the scene consists of a single plane due to the significant distance of the scene from the camera. Hence, W_m can be obtained by using affine SIFT [13] descriptors to match corresponding points in the different frames. However, sometimes scenes consist of multiple planar regions (apart from the occlusions) for which pixel motions are different. Furthermore, scene elements can be dynamic leading to non-global pixel motion. We also address de-fencing of such scenes in this work. Again the depth data from Kinect enables us to segment out the various planar regions at different depths from the sensor. ASIFT descriptors are then used to find pixel motion corresponding to the individual segments. For some challenging cases we have also used the method in [14] to obtain dense optical flow. This is particularly useful if the scene contains dynamic objects.

The maximum a-posteriori estimate of the de-fenced image can be obtained as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{y}_m - O_m W_m \mathbf{x}||^2 + \beta \sum_{c \in C} V_c(\mathbf{x}) \tag{2}$$

where β is the regularization parameter. The joint pdf of the MRF can be specified as Gibbsian by the Hammersley-Clifford theorem [18]

$$P(\mathbf{x}) = \frac{1}{Z} \exp(-V_c(\mathbf{x})) \tag{3}$$

where Z is the partition function and $V_c(\mathbf{x})$ is the clique potential function. We choose a robust form for the clique potential function as $V_c(\mathbf{x}) = |x_{i,j} - x_{i,j+1}| + |x_{i,j} - x_{i,j-1}| + |x_{i,j} - x_{i-1,j}| + |x_{i,j} - x_{i+1,j}|$ considering a first-order neighbourhood. We minimize the objective function in Eq. (2) by using the loopy belief propagation (LBP) technique [17]. The parameter β is chosen as 5×10^{-4} for all our experiments.

Analogous to the procedure for image de-fencing, we relate the occluded 'fenced' depth maps recorded by Kinect to the inpainted depth profile.

$$\mathbf{d}_m = O_m W_m \overline{\mathbf{d}} + \mathbf{n}_m \tag{4}$$

where $\mathbf{d}_m$ represents the occluded depth maps due to the fence and $\overline{\mathbf{d}}$ denotes the completed depth profile with the explanation of other symbols being identical to Equation (1). The maximum a-posteriori estimate of the completed depth profile can be obtained as

$$\hat{\overline{\mathbf{d}}} = \arg\min_{\overline{\mathbf{d}}} ||\mathbf{d}_m - O_m W_m \overline{\mathbf{d}}||^2 + \alpha \sum_{c \in C} V_c(\overline{\mathbf{d}}) \tag{5}$$

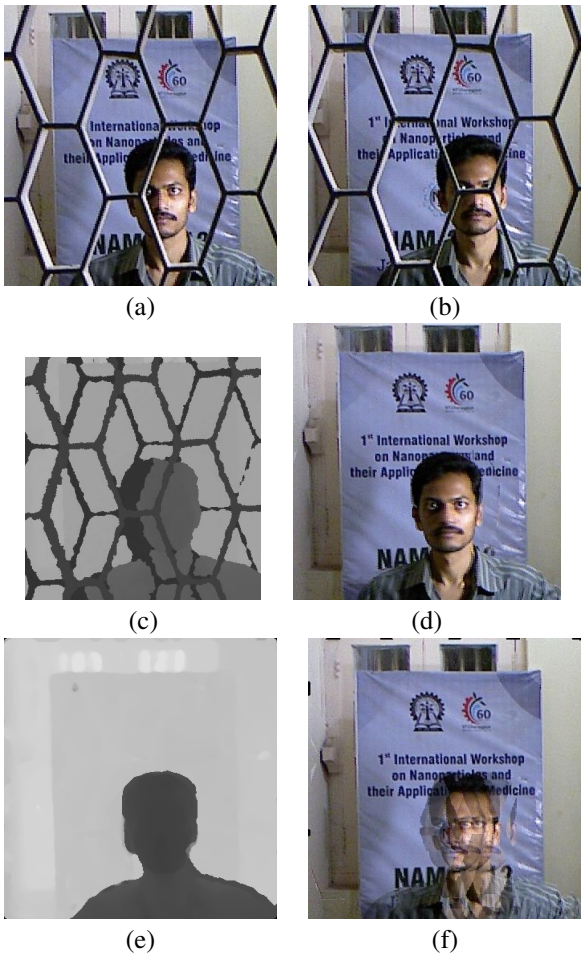
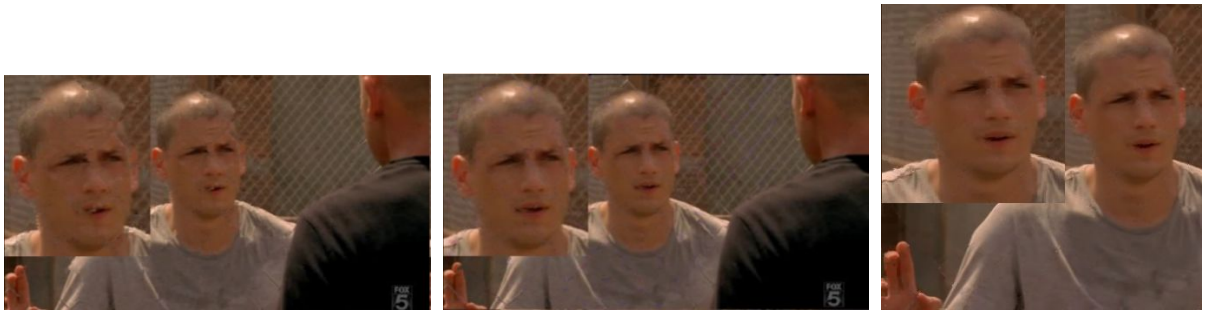


Fig. 4. (a), (b) First and third observations. (c) Depth map corresponding to the first observation. (d) De-fenced image obtained using the proposed method. (e) Depth completion using the proposed method. (f) Output of method in [1].

where α is the regularization parameter. Since we have a-priori aligned the depth profiles recorded by Kinect with the corresponding RGB images, we can use the estimates of relative motion between the color frames obtained earlier to align the multiple captured depth maps. Therefore, the matrices W_m and O_m in Eq. 5 remain identical to those in Equation 1. Similar to the estimation of the de-fenced image, we used a robust form for the clique potential function, $V_c(\overline{\mathbf{d}}) = |\overline{d}_{i,j} - \overline{d}_{i,j+1}| + |\overline{d}_{i,j} - \overline{d}_{i,j-1}| + |\overline{d}_{i,j} - \overline{d}_{i-1,j}| + |\overline{d}_{i,j} - \overline{d}_{i+1,j}|$ considering a first-order MRF neighbourhood. We minimize the objective function in Eq. (5) by using the loopy belief propagation (LBP) technique [17]. The parameter α is chosen as 5×10^{-4} for all our experiments. We note that the proposed technique is robust to changes in the values of the smoothness parameters β and α. It is to be observed that the depth maps recorded using Kinect are prone to artifacts/holes due to shadows/occlusions etc. Since we use multiple recorded depth profiles using the Kinect sensor, it is highly unlikely that depth data is missing at a particular spatial location in all the depth maps. Therefore, our optimization-based framework is able to inpaint the depth map effectively.



(a) (b) (c)

Fig. 5. (a) De-fenced image obtained using the method in [2]. (b) Output of [3]. (c) De-fenced image obtained using the proposed method.

III. Experimental results

Initially, we report experiments wherein the relative pixel displacements are assumed to be global. We actually capture a video by panning the Kinect sensor which records both the RGB data and the corresponding depth data. In Fig. 1 (a), we show a painting occluded by a fence. We show the depth map for the first observation in Fig. 1 (b). Note that we can easily segment out the fence using this depth data by applying Otsu's method since the painting in the background is at a different distance from the sensor. A robust estimate of the segmented fence mask can be obtained after applying image dilation operation to fence pixels and this is shown in Fig. 1 (d). We use the affine SIFT [13] descriptor to obtain the relative global pixel shifts between 4 color observations chosen from the captured video. The basic idea behind the choice of a particular set of four observations is that data (image or depth) which is occluded in one frame should be revealed in the other images. If we are able to satisfy this requirement to a reasonable degree the data term in Eq. 2 guides the image de-fencing procedure well. Sample corresponding point matches between the first and second frames are depicted in Fig. 1 (c). The estimated pixel shifts between the four captured frames were (-0.87, 4.42), (-2.9, 3.46), (-2.23, 8.45) relative to the first observation. The de-fenced image estimated using the proposed method is shown in Fig. 1 (e). It is clear that the fence occlusions have been very effectively removed. Similarly it is possible to obtain a good estimate of the "de-fenced" depth map although we do not show it here for brevity since depth profile is merely planar.

Next, we show results obtained with a video wherein the frames contained non-global pixel motion. In Figs. 2 (a), (b) we show two frames from such a video. Since optical flow algorithms [12] do not give accurate estimates of the pixel motion, this is a challenging scenario for de-fencing. The depth map recorded by the Kinect sensor corresponding to the first observation is shown in Fig. 2 (c). It is easy to observe that the fence occlusion locations can be obtained by using Otsu's method for segmentation. We also observe from the depth map that the scene consists of predominantly two depths. Therefore, we assume that the scene is composed of two planar regions at different depths from the Kinect. The segmented foreground region is given in Fig. 2 (d). To obtain the relative displacements between the foreground and the background regions, respectively, in the four frames chosen from the captured video, we use ASIFT descriptors. Sample matches of corresponding points in the first and second observations are shown in Fig. 2 (e). The pink lines connect corresponding points in the background and the white lines connect points in the foreground. The estimated pixel motion between frames for the foreground region is (-2.74, -6.1), (-9.6, -2.95) and (-12.62, -0.3), respectively. The pixel shifts for the background region were estimated as (-1.24, -0.84), (0.49, 1.14) and (1.32, -0.94). The de-fenced image obtained using the proposed algorithm is shown in Fig. 3 (a). The effectiveness of our approach for even such a case involving complex motion is clearly evident. In Fig. 3 (b), we show the completed depth profile corresponding to the first observation. Since there are only two depth layers at a significant distance from the Kinect sensor, we are able to successfully estimate them by assuming them to be planar.

As a challenging example, we show results for another video wherein the relative displacements are non-global. In Fig. 4 (a), (b) we show the first and third observations obtained from a video captured by panning the Kinect camera. Note that the fence pattern here is thicker in width and different from that in Fig. 2. We used only three frames from the captured video for this case. The estimated pixel shifts for the foreground region is (2, 19) and (4, -12) relative to the first frame. The shifts for the background region were obtained as (2.5, 3.73) and (4.8, 12.1), respectively. The depth map obtained using Kinect is shown in Fig. 4 (c). The de-fenced image corresponding to the first frame is shown in Fig. 4 (d). Observe that there are hardly any artifacts and all edges are faithfully reconstructed. It is interesting to observe the estimated depth profile of the scene using our method shown in Fig. 4 (e). Initially, we point out that there are multiple depth layers in the scene. The Kinect captured depth map corresponding to the first observation, depicted in Fig. 4 (c), has several artifacts due to shadowing effects and occlusions by the fence. We have used Otsu's method to obtain the fence masks from each of the three captured depth profiles. As shown in Fig. 4 (e), we can easily discern three major depth layers corresponding to the person, the poster and the window in the background. It is interesting to observe that it is possible to even reconstruct the depth of the rods in the window just above the upper edge of the poster in the background.

We compare our method with the approach proposed in [1]. The technique in [1] makes the assumption that the frames of the captured video are relatively shifted by global pixel motion. Hence, it cannot handle the scenario wherein there are multiple depths in the scene to be de-fenced. We give as input to [1] the motion of the background pixels in the 3 frames as the global motion and the output is shown in Fig. 4 (f). As

expected, the background region is de-fenced to a reasonable degree but there are ghosting effects in the foreground region due to wrong pixel motion. The other major drawback of the technique in [1] is that the fence pixels have to be marked manually by a user and hence the method is not automatic. Finally, the technique of [1] is only restricted to image de-fencing using RGB videos and does not address the problem of inpainting the depth profile of the scene.

We compare our technique with previous works for image de-fencing [2], [3]. In Fig. 5 (a), we show one of the results of [2]. We can clearly observe several artifacts on the right side of the head of the person facing the camera. Also, the lips have not been reconstructed properly. Particularly, note the close-up shown on the left side of the figure. In [2], only a single image has been used for removal of the fence occlusions. However, the authors of [3] have used a video for the same example. Using multiple frames they attempt to de-fence the image and the result is shown in Fig. 5 (b). In the close-up shown in the inset of Fig. 5 (b), we can see that there are still some artifacts near the lip region although the result is better as compared to Fig. 5 (a). We used the same example video and extracted four frames from it. Note that we do not have depth data for this example since the video was not captured using the Kinect sensor. We have used the method in [19] to isolate the fence masks from the four observations. The optical flow between the frames was estimated using the method in [14]. The de-fenced image using our algorithm is shown in Fig. 5 (c). We can clearly observe the improvement in the reconstruction of the de-fenced image over the outputs of [2] and [3]. There are hardly any artifacts and the face is properly reconstructed.

In the next experimental result we use complex shapes and multiple depths to demonstrate the effectiveness of the proposed algorithm. In Figs. 6 (a) and (b), we show the first and third observations obtained from the captured RGB video. There are multiple scene elements including a star-shaped object. The face of the person is heavily occluded with a thick fence. In Fig. 6 (c), we show the depth profile corresponding to the first observation affected by fence occlusions. Observe that there are several holes in this depth map due to shadowing and occlusion effects. The edges of the "star" shaped are affected adversely by occlusions. Since, in this case, the scene consists of several subtle depth variations, we obtain the relative motion between the frames using a recently proposed approach for estimating dense optical flow [14]. The de-fenced image is shown in Fig. 6 (d). The effectiveness of the proposed technique is evident as seen from the accurately estimated boundaries of the different objects in the scene. Also, the thick occlusions on the face are completely removed. The optical flow between the first and third frames is shown in Fig. 6 (e). The completed depth profile is depicted in Fig. 6 (f). Unlike the scenario of image de-fencing, for the case of depth inpainting, we have to treat the artifacts in the depth maps due to shadows also as occlusions since depth data is missing at those spatial locations. Hence, depth inpainting is a more challenging problem as compared to image de-fencing. Also, this result shows that our technique is able to handle complex motion of scene pixels since there are variations of depth values in the various objects comprising the scene. Particularly, note the variations of depth on the body of the person and the ramp-shaped depth profile of the calendar.

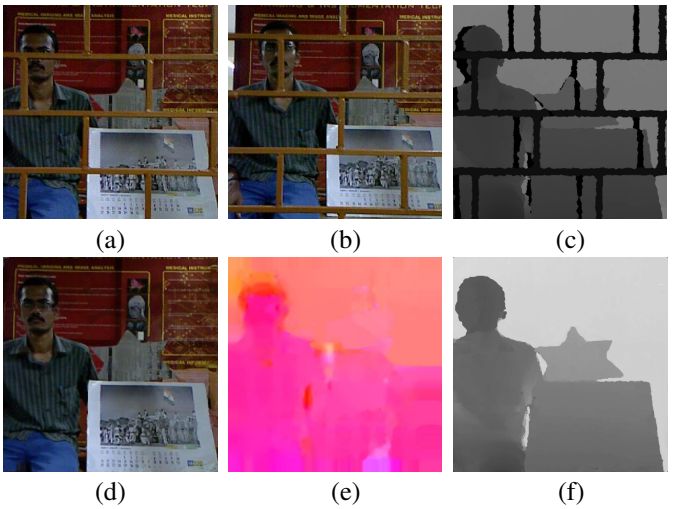


Fig. 6. (a), (b) First and third observations from the captured color video. (c) Occluded depth profile obtained using Kinect corresponding to the first observation. (d) De-fenced image using the proposed method. (e) Optical flow between frames 1 and 3 using the technique in [14]. (f) Inpainted depth map using our method.

Until now we have shown the results obtained for the scenario wherein the scene elements are static and the Kinect sensor is moved while capturing a video. We now consider a case where the scene is also dynamic in nature. Specifically, we captured data when a person is walking in a corridor behind a fence. We show the first and fourth observations from the captured color video in Figs. 7 (a) and (b). The depth profile corresponding to the first frame is shown in Fig. 7 (c). Again, we estimated the relative displacements between the observations using [14]. The de-fenced image is shown in Fig. 7 (d). The accuracy of image reconstruction is high in spite of multiple depth values on the body of the individual. The optical flow between the first and the fourth frames is shown in Fig. 7 (e). The inpainted depth map is shown in Fig. 7 (f), wherein, the depth variations over body of the person and the accuracy of reconstruction of the depth boundaries can be easily observed.

Finally, we show the performance of the proposed method for yet another challenging case wherein we consider complex motion of scene elements. In Figs. 8 (a) and (b), we show the first and the third observations, repectively from a video. Here a person behind a fence is moving his head by a significant amount. Note that the fence is different and much thicker than the cases considered in Figs. 2, 6 and 7. The occluded depth profile corresponding to the first observation is shown in Fig. 8 (c). Observe the significant number of spatial locations wherein depth data of the background is corrupted or missing. We used four observations from the color video to obtain the de-fenced image shown in Fig. 8 (d). We note that the vertical edge of the wall which is missing in the first and third observations shown in Figs. 8 (a) and (b) is properly reconstructed in Fig. 8 (d). We obtain the relative motion between the frames using a recently proposed approach for estimating dense optical flow [14]. The optical flow between the first and second frames is shown in Fig. 8 (e). The effectiveness of the proposed technique is adequately demonstrated in the accurate reconstruction of the depth map in Fig. 8 (f), where the background scene consists of several depth layers and fine variations.

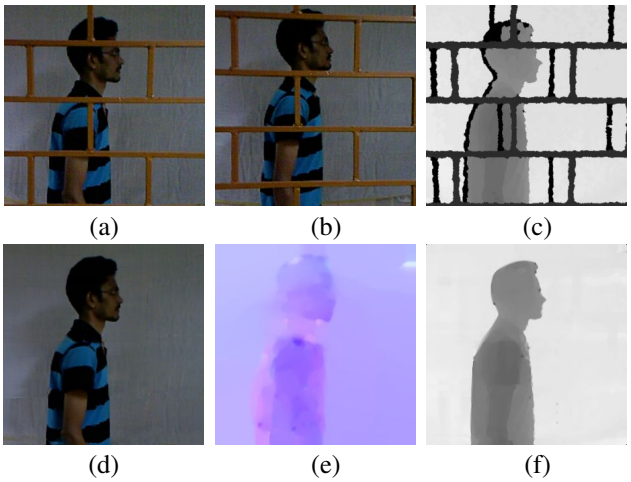


Fig. 7. (a), (b) First and fourth observations from the captured color video with a person walking behind a fence. (c) Occluded depth profile obtained using Kinect corresponding to the first observation. (d) De-fenced image using the proposed method. (e) Optical flow between frames 1 and 4 using the technique in [14]. (f) Inpainted depth map using our method.

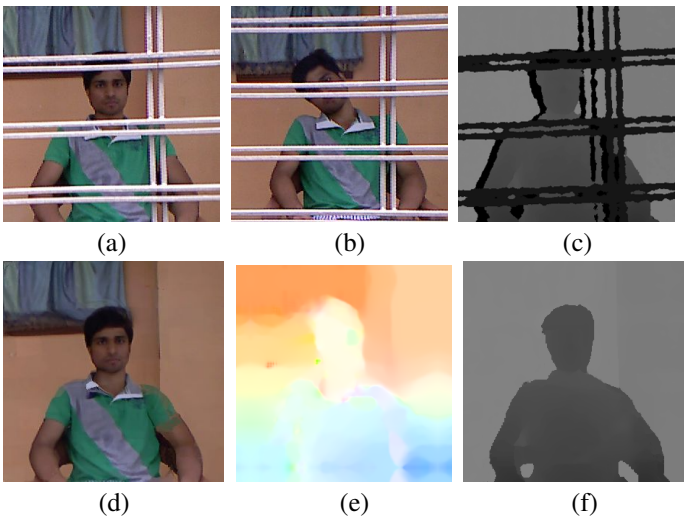


Fig. 8. (a), (b) First and third observations from the captured color video with a person moving his head. (c) Occluded depth profile obtained using Kinect corresponding to the first observation. (d) De-fenced image using the proposed method. (e) Optical flow between frames 1 and 2 using the technique in [14]. (f) Inpainted depth map using our method.

We provide the input data and results obtained for each experiment in the supplementary data accompanying our paper.

IV. Conclusion

We proposed a multi-modal approach for "de-fencing" a scene using RGB-D data. Specifically, we use multiple frames from a video and the aligned depth maps captured by panning the scene with the Microsoft Kinect sensor. We addressed the problem of identification of the fence pixels by using the captured depth data. We considered several scenarios regarding the motion of pixels in the frames of the video. Firstly, we assumed global motion of all scene pixels considering the significant distance of the scene from the sensor. Next, we also addressed the challenging scenario of non-global motion as well as dynamic scene elements. We proposed an optimization-based approach by solving for the maximum a-posteriori estimate of the de-fenced image and the inpainted depth profile assumed to be two distinct Markov random fields. Our results show the effectiveness of the proposed algorithm for real-world data.

As part of future work, we will develop a real-time automatic image de-fencing and depth completion algorithm which will be useful with the advent of cameras equipped with depth sensors.

References

[1] V. Khasare, R. Sahay, and M. Kankanhalli, "Seeing through the fence: Image de-fencing using a video sequence," *20th IEEE International Conference on Image Processing (ICIP)*, pp. 1351–1355, 2013.

[2] Y. Liu, T. Belkina, J. Hays, and R. Lublinerman, "Image de-fencing," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[3] M. Park, K. Brocklehurst, R. T. Collins, and Y. Liu, "Image de-fencing revisited," *Asian Conference on Computer vision*, pp. 422–434, 2011.

[4] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting : Joint color and depth completion from stereo images," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[5] A. V. Bhavsar, and A. .Rajagopalan, "Inpainting large missing regions in range images," *International Conference on Pattern Recognition (ICPR)*, pp. 3464–3467, August 2010.

[6] R. R.Sahay and A. N. Rajagopalan, "Joint image and depth completion in shape-from-focus: Taking a cue from parallax," *Journal of the Optical Society of America A (JOSA A)* , vol. 27, no. 5, pp. 203–1213, 2010.

[7] A. V. Bhavsar, and A. N. Rajagopalan, "Range map super-resolution inpainting and reconstruction from sparse data," *Computer Vision and Image Understanding (CVIU)*, vol. 116, no. 4, pp. 572–591, 2012.

[8] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles, "Image inpainting," *Proc. ACM SIGGRAPH*, pp. 417–424, July 2000.

[9] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 882–889, Aug. 2003.

[10] A. Criminisi, P. Prez, and K. Toyama, "Region filling and object removal by exemplar-based inpainting," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, Sept. 2004.

[11] A. Bugeau, M. Bertalmio, V. Caselles, and G. Sapiro, "A comprehensive framework for image inpainting," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2634–2645, Oct. 2010.

[12] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2432–2439, 2010

[13] J. M. Morel and G.Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[14] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, no. 3, pp. 500–513, 2011.

[15] M. Park, K. Brocklehurst, R. Collins, and Y. Liu, "Deformed lattice detection in real-world images using mean-shift belief propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 31, no. 10, pp. 1804–1816, 2009.

[16] R. C Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2008.

[17] D. P. Huttenlocher and P. F. Felzenszwalb., "Efficient belief propagation for early vision," *International Journal of Computer Vision,*, vol. 70, pp. 41–54, 2006.

[18] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Springer, 2001.

[19] Y. Zheng, and C. Kambhamettu, "Learning based digital matting," *IEEE International Conference on Computer Vision*, pp. 889–896, 2009.