2011 International Conference on Computer Science and Information Technology (ICCSIT 2011)
IPCSIT vol. 51 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V51.40

Super-resolution for Depth Maps

Cihui Yang[1+], Hong Liu[1], Richard Green[2], Enmin Song[1] and Meng Fan[1]

[1]School of Computer Science and Technology, Huazhong University of Science and Technology, Key Laboratory of Education Ministry for, Image Processing and Intelligent Control, Wuhan, China

[2]The Computer Science department, University of Canterbury, Christchurch, New Zealand

Abstract. Although time of flight (TOF) cameras are becoming more popular, their low spatial resolution limits their usefulness. In this paper, we propose two new super-resolution methods to improve the resolution of depth maps generated by TOF cameras. One is based on the Lucas Kanade (LK) optical flow algorithm, and another is based on the scale-invariant feature transform (SIFT) algorithm. The results show that the proposed LK optical flow based method is more accurate and efficient in improving the spatial and depth resolution of depth maps than the proposed SIFT based method. It is more useful for real-time processing of depth images.

Keywords-Super-resolution; depth map; Lucas Kanade optical flow; SIFT

1. Introduction

Depth imaging is a very useful technology in computer vision. It has already been used in many applications, such as human interfaces, manufacturing, robotics, security and games, etc. Existing depth sensing methods can be categorized into three classes: Laser range scanners, Stereo vision, and Time-of-Flight (TOF) imaging. TOF imaging captured the depth information by emitting pulses of infra-red light to all objects in the scene and measuring the return travel time of the reflected light [1]. It is independent of the scene and can provide real-time and accurate depth information. So TOF imaging has recently became more popular, but the resolution of this imaging is still very limited.

Super-resolution is a technology to obtain a high resolution (HR) image from one or multiple low resolution (LR) images of the same scene. It has been developed over the last two decades since its first proposed in 1984 by Tsai and Huang [2]. Most of the methods researched are for images or video. Only recently have super-resolution approaches [3-9] been proposed for depth map to improve the resolution of depth data. In this paper, we present two new super-resolution methods to improve the resolution of depth maps. One is based on the Lucas Kanade (LK) optical flow algorithm, and the other is based on the scale-invariant feature transform (SIFT) algorithm. The principle idea behind the proposed methods is obtaining depth maps of a scene from slightly displaced viewpoints and then aligning them, whereby they are fused together to generate a HR depth map. The LK optical flow algorithm and SIFT algorithm perform registration to support this alignment. Both the methods are computationally efficient and proved to be useful for enhancing the spatial (x, y direction) and depth (z direction) resolution of depth maps.

2. Related Work

Depth super-resolution is relatively new in computer vision. Only a few super-resolution methods for depth maps have been proposed so far. One approach to improve the resolution of depth maps is to integrate color images taken from a viewpoint close to that of the depth sensor [3-4]. Although these methods are often computationally efficient, having to use two cameras constrains potential applications. Another approach to

+ *E-mail address*: yangcihui@smail.hust.edu.cn

enhance depth resolution is to integrate multiple LR depth maps, which differs from the previous method in that only depth maps are now used [5-6].

Some researchers also tried to apply image super-resolution algorithms to depth maps. Rosenbush et al. [7] adopted frequency-domain method to improve the resolution of range data and showed that ideas from image super-resolution can be applied to range map. Schuon et al. [8] applied the image super-resolution method proposed by Farsiu et al. [10] to depth maps. They [9] also presented a depth super-resolution approach in which they use an energy minimization framework to make the super-resolution result be coherent with the LR depth maps. We also adopted the idea but used different image super-resolution algorithm.

3. Our Algorithm

In the process of obtaining a depth map of a scene by a depth sensor, there is a natural loss of spatial resolution because of insufficient sensor density. Gaussian noise and motion blur are also introduced due to low IR illumination levels. To break through the resolution limit of the sensor and obtain more depth information, we can generate different depth measures of the scene by moving the sensor or the scene. Then, every depth map contains some depth information of the same scene which is different from other depth maps. By fusing these depth maps, a HR depth map of the scene can be generated.

It is assumed that N depth maps of the same scene have been captured. These depth maps can be viewed as different representations of a true HR depth map. We assume that the formation process of a depth map is similar to the formation process of a normal image. Then, we employ the following model proposed by Elad and Feuer [12] to represent the relationship between the measured LR depth map $\mathbf{Y}_k$ $(1 \le k \le N)$ and the ideal HR depth map $\mathbf{X}$.

$$\mathbf{Y}_k = \mathbf{D}_k \mathbf{B}_k \mathbf{M}_k \mathbf{X} + \mathbf{n}_k, \quad 1 \le k \le N, \tag{1}$$

where $\mathbf{M}_k$ is a warp matrix denoting the geometric warp executed on depth map $\mathbf{X}$. $\mathbf{B}_k$ represents a blur matrix. $\mathbf{D}_k$ represents a down-sampling matrix. $\mathbf{n}_k$ is a noise vector.

Our algorithm is based on the approach proposed by Chiang and Boult [11]. It includes the following steps: First, pre-process all the LR depth maps and remove the noise. Then, select one depth map from the given LR depth maps as reference depth map and estimate sub-pixel motion between the remained depth maps with the reference depth map. In this step, either the SIFT or LK optical flow method is used to do registration. After motion estimation, all the pixels of the LR depth map are mapped to corresponding position on a HR depth map. The forth step is to reconstruct the HR depth map by fusing all the warped depth data together. Then, noise removal results in the final HR depth map.

3.1. Preprocessing

There exists lots of Gaussian noise in the depth maps captured with TOF cameras, which significantly degrades the quality of these depth maps. It is important to avoid propagating data from the noise during the reconstruction process. So, the noise should be removed first. Considering the characteristics of depth maps captured by TOF cameras, we adopt threshold method to filter invalid depth values according to the distance between the camera and the scene.

3.2. Sub-pixel motion estimation

For each pixel in LR depth maps, it is necessary to know the corresponding point in the desired HR depth map. Thus, motion estimation is performed in this step to calculate the mapping relationship. We first select one reference depth map, and then calculate the mapping relationship between the other depth maps and the reference depth map. For a point $[s\ t]^T$ in depth map $\mathbf{Y}_k$ $(1 \le k \le N)$, its corresponding point $[u\ v]^T$ in the reference depth map can be represented as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} r_{1,k} & r_{2,k} \\ r_{3,k} & r_{4,k} \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} + \begin{bmatrix} m_{1,k} \\ m_{2,k} \end{bmatrix}, \tag{2}$$

where $r_{1,k} \sim r_{4,k}$ represent rotation, scale and stretch. $m_{1,k}$ and $m_{2,k}$ represent translation. So, the goal of this step is to calculate $r_{1,k} \sim r_{4,k}$, $m_{1,k}$ and $m_{2,k}$.

Some researchers applied the LK optical flow algorithm or SIFT algorithm into super-resolution for registration respectively [9, 13-14]. Most of the methods are proposed for images. Only Schuon et al. [9] apply the LK optical flow algorithm into depth maps. In this paper, we apply the LK optical flow algorithm and SIFT algorithm into depth map super-resolution, but use different super-resolution and reconstruction method from Schuon et al. [9].

1) Registering by the LK optical flow algorithm

Optical flow is a velocity field of an image [15]. It is well suited for measuring small changes in the global transformation and can reach sub-pixel accuracy. The key idea of optical flow algorithm is that, for each point (u,v) in frame t, there will be a corresponding point in frame $t+\Delta t$. Formally, if $I(u(t),v(t),t)$ denotes the continuous space-time intensity function of an image, then

$$I(u(t),v(t),t) \approx I(u(t+\Delta t),v(t+\Delta t),(t+\Delta t)). \tag{3}$$

Apply Taylor expansions to the right side of (3), we can get

$$\frac{\partial I}{\partial u}\frac{du}{dt}+\frac{\partial I}{\partial v}\frac{du}{dt}+\frac{\partial I}{\partial t}=0, \tag{4}$$

where $\frac{du}{dt}$ and $\frac{du}{dt}$ are the components of the image velocity vector. $\frac{\partial I}{\partial u}$ and $\frac{\partial I}{\partial v}$ are the components of spatial gradient ∇I. $\frac{\partial I}{\partial t}$ is partial differentiation of I with respect to time. Lucas Kanade (LK) [16] is one of the techniques to solve (4). It introduces an additional term by assuming the flow to be constant in a local neighborhood around the central pixel.

When using the LK optical flow method to estimate motion, Harris corners are detected first in the reference depth map. Then, the corresponding points in the other depth map are calculated using the LK optical flow method.

2) Registering by the SIFT algorithm

Scale invariant feature transform (SIFT) [17] is an algorithm to detect and describe local features in images. The features extracted by SIFT are invariant to image scale, rotation and illumination, and partially invariant to affine transformation. The features are also highly distinctive. SIFT is also computationally efficient, so that two images can be matched with near real-time performance. These advantages make it suitable for real-time registration of LR depth maps.

After finding the SIFT features in two depth maps, the best match of a keypoint in one depth map can be found by identifying the keypoint with minimum Euclidean distance for the feature vector in another depth map.

3) Motion estimation

Assuming that *l* pairs of matching points are generated in step 2, (2) can be rewritten as

$$\begin{bmatrix} s_1 & t_1 & 0 & 0 & 1 & 0 \\ 0 & 0 & s_1 & t_1 & 0 & 1 \\ s_2 & t_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & s_2 & t_2 & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \\ s_l & t_l & 0 & 0 & 1 & 0 \\ 0 & 0 & s_l & t_l & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,k} \\ r_{2,k} \\ r_{3,k} \\ r_{4,k} \\ m_{1,k} \\ m_{2,k} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ \vdots \\ u_l \\ v_l \end{bmatrix} \tag{5}$$

where $(s_i,t_i),(u_i,v_i)$ $(i=1,2,\cdots,l)$ denotes the matching points in the two depth maps. The motion parameters can be solved by the least squares method.

3.3. Warping

After motion estimation, warping is now done in this step. For a point $[s\ t]^T$ in LR depth map $\mathbf{Y}_k$ $(1 \le k \le N)$, its corresponding point $[u_{warp}\ \ v_{warp}]^T$ in the estimated depth map $\mathbf{Y}_k'$ can be computed as:

$$\begin{bmatrix} u_{warp} \\ v_{warp} \end{bmatrix} = \beta * \left(\begin{bmatrix} r_{1,k} & r_{2,k} \\ r_{3,k} & r_{4,k} \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} + \begin{bmatrix} m_{1,k} \\ m_{2,k} \end{bmatrix} \right) \tag{6}$$

For the points of depth map $\mathbf{Y}_k'$, which have no corresponding points in the depth map $\mathbf{Y}_k$, their depth value are assigned with -1. Thus, multiple estimations of an underlying HR depth map can be obtained.

3.4. Super-resolution reconstruction

On the assumption that the resolution of an original LR depth map is $m \times n$, we gather all the estimated HR depth maps together and describe it as

$$\mathbf{R} = \begin{bmatrix} \mathbf{a}_{1,1} & \cdots & \mathbf{a}_{1,\beta n} \\ \vdots & \ddots & \vdots \\ \mathbf{a}_{\beta m,1} & \cdots & \mathbf{a}_{\beta m,\beta n} \end{bmatrix} \tag{7}$$

where $\mathbf{a}_{i,j} (i = 1,2,\cdots,\beta m;\ j = 1,2,\cdots,\beta n)$ denotes the vector that consist of the pixel values of the N depth maps in location (i, j). When generating $\mathbf{a}_{i,j}$, the value -1 will be discarded. By processing $\mathbf{R}$, we can get the approximate estimation of a HR depth map. For $\mathbf{a}_{i,j}$ which is not empty, all its elements are sort first. Then some of the biggest values and smallest values are removed. By calculating the mean of the remaining values, the approximate value of HR depth map in location (i, j) can be obtained. After processing all the nonempty elements of $\mathbf{R}$, the approximate values of the pixels that have no values in the LR depth map can be found by calculating the mean of their neighborhood.

3.5. Removing the noise

There is always some noise in the resulting depth map, which is brought by the original LR depth maps. In this step, we remove some noise by threshold method first, then apply a median filter to remove the remaining salt-and-pepper noise.

4. Experimental Results

The experiments were performed with a Canesta camera, which generates a depth map of 64×64 pixels resolution. We select the up-sampling multiple as 8. A *snowdrop* plant and a *dieffenbachia* plant are selected as experimental objects. When obtaining the depth maps, the Canasta camera is put in front of the object, and the depth maps are captured by translating and rotating the camera. 16 LR depth maps are selected to generate a HR depth map. Fig. 1 shows the color images of the snowdrop plant and the dieffenbachia plant.

(a) Snowdrop

(b) Dieffenbachia

Figure 1. Color images of snowdrop and dieffenbachia

In the first experiment, the distance between the camera and the snowdrop is about 45 cm. The shutter time of the camera has been set to suppress noise and limit the depth range, so only the snowdrop can be seen in the original LR depth map. Fig. 2(a) shows one of the original LR depth maps. The up-sampled depth map by linear interpolation is showed in Fig. 2(b). Fig. 2(c) and Fig. 2(d) show the super-resolution results generated by our

method based on the SIFT algorithm and LK optical flow algorithm respectively. The corresponding pseudo-color images of these depth maps are showed in the second row of Fig. 2. In Fig. 2(a) and Fig. 2(b), we can see obvious LR saw-tooth edge. While in Fig. 2(c) the HR edge is relatively smooth, and some detail information has been recovered. However, the detail information is not enough. In Fig. 2(d), the edge is smoother than the first three ones, and more detail has been recovered. For example, the leftmost and the rightmost leaves in Fig. 2(a) and Fig. 2(b) are not clear and smooth. In Fig. 2(c), the same leaves are smooth but not very clear, and some information of the rightmost leaf has been lost. By contrast, in Fig. 2(d), the leaves are very smooth and clear. Another example is the rightmost two flowers. We cannot recognize them in Fig. 2(a) and Fig. 2(b), whereas they are clear and easily recognized in Fig. 2(d). In Fig. 2(c), we can also recognize the flowers, but they are not very clear compared with Fig. 2(d).

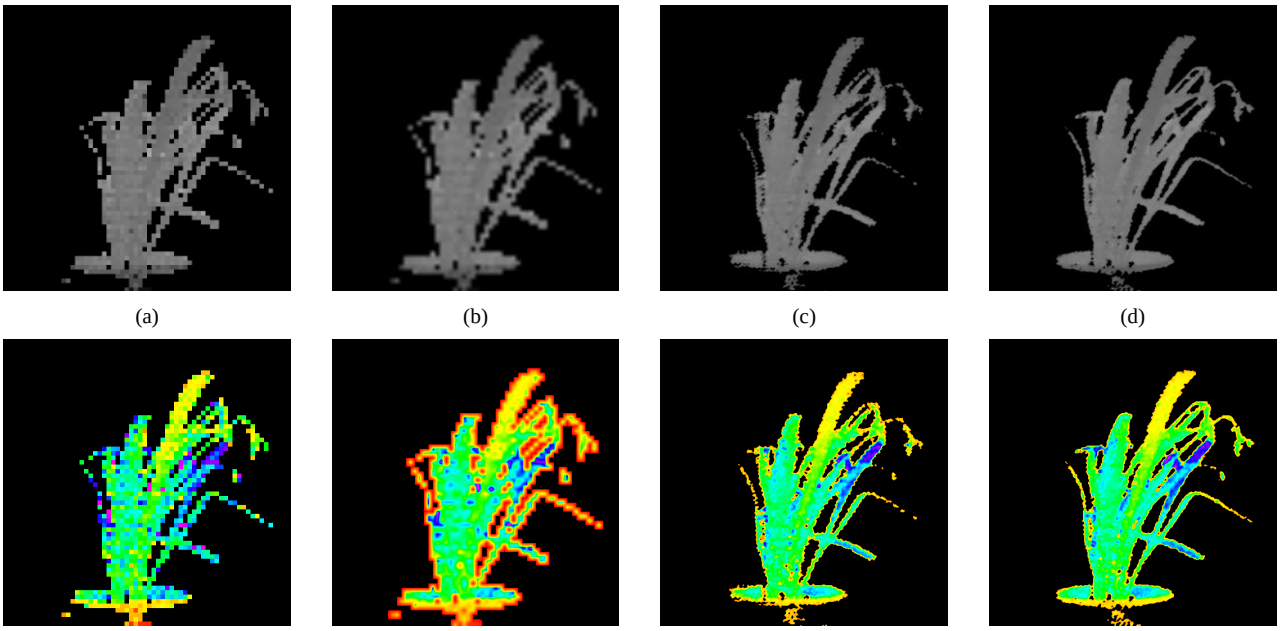


Figure 2. Super-resolution result of snowdrop's depth map: The first row depicts depth maps, and the second row depicts pseudo-color images of these depth maps; (a) The 1st original LR depth map; (b) Up-sampled depth map by linear interpolation; (c) Super-resolution result by SIFT based method; (d) Super-resolution result by LK optical flow based method; (e)-(h) Pseudo-color images of (a)-(d).

As well as the improvement in spatial (x, y direction) resolution, we can also see depth (z direction) resolution enhancement from the first experiment's results. In the pseudo-color image of the original LR depth map in Fig. 2(e), there are some discrete red points which represent larger depth value than green and blue points inside the leaves, and the change from one color to another color is very sharp. There are also some discrete red points in the pseudo-color image of up-sampled depth map in Fig. 2(f). While in the pseudo-color images of super-resolution results in Fig. 2(g) and Fig. 2(h), the red points are replaced by correct green points, and the color change is very gradual.

In experiment 2, the camera is approximately 40 cm away from the dieffenbachia. Fig. 3 shows the results from experiment 2. From these results, we can also see an obvious saw-tooth edge in the original LR depth map and the up-sampled depth map by linear interpolation in Fig. 3(a) and Fig. 3(b) respectively. The edges are smooth in the super-resolution result depth maps generated by the proposed methods in Fig. 3(c) and Fig. 3(d). In Fig. 3(c), some holes can be seen in the right-top two leaves, and yet they are not in Fig. 3(d). Similar to experiment 1, a gradual depth change can also be seen in Fig. 3(g) and Fig. 3(h), while the change is quick in Fig. 3(e) and Fig. 3(f).

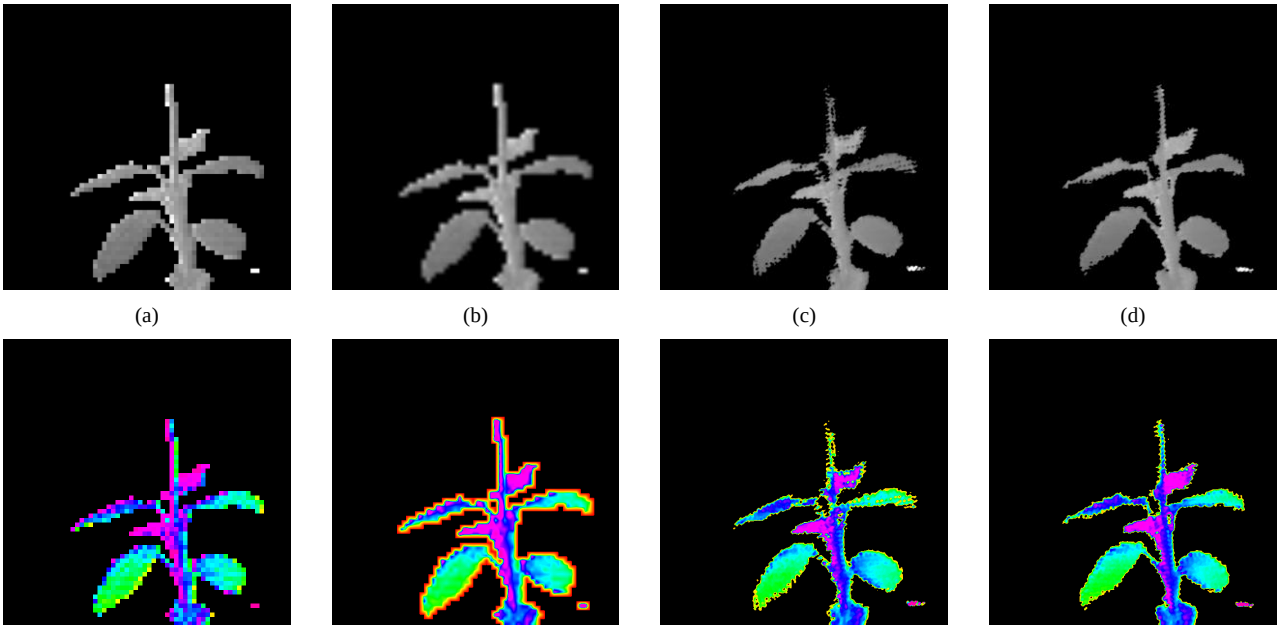
This experiments shows that the proposed methods both prove to be useful for improving the spatial and depth resolution of depth maps, where the LK optical flow based method is better than the SIFT based method. In fact, although for small depth maps, some SIFT features can be obtained. The registration by the SIFT method is still not very robust, especially for the depth maps with minimal depth texture. The LK optical flow algorithm

is good at measuring small changes in the global transformation, even for small depth maps. Thus, the LK optical flow based method achieves better performance over the SIFT method.

Table I shows elapsed times of the super-resolution reconstruction processes for the above experiments. The experiments are executed on a PC with an Intel Core 2 Duo 2.66GHz CPU (2MB×2 L2 Cache) and 2GB RAM. This table shows that the LK optical flow based method can finish the whole super-resolution construction in 1.6 seconds which is faster than the SIFT based method.

TABLE I Comparison of experiment time Unit: s

Method / Experiment	LK optical flow based method	SIFT based method
Snowdrop	1.55	4.50
Dieffenbachia experiment	1.51	4.45
Speaker experiment	1.55	3.12



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3. Super-resolution result of dieffenbachia's depth map: The first row depicts depth maps, and the second row depicts pseudo-color images of these depth maps; (a) The 2nd original LR depth map; (b) Up-sampled depth map by linear interpolation; (c) Super-resolution result by SIFT based method; (d) Super-resolution result by LK optical flow based method; (e)-(h) Pseudo-color images of (a)-(d).

5. Conclusion

In this paper, we present two methods to enhance the resolution of depth maps. One is based on the SIFT algorithm, and another is based on the LK optical flow algorithm. Experimental results show that the LK optical flow based method is better at handling depth maps with low depth texture than the SIFT based method, and can improve the spatial and depth resolution at the same time. For depth maps with a more depth texture, both methods demonstrate good results in enhancing the spatial and depth resolution, but the LK optical flow based method is better and faster than the SIFT based method. The proposed LK super-resolution depth map algorithm is much more efficient and so is more useful for real-time processing of depth images.

6. Acknowledgment

This work was supported in part by the China International Science and Technology Cooperation Project (Grant No. 2009DFA12290).

7. References

[1] G. J. Iddan and G. Yahav, "3D imaging in the studio (and elsewhere ...)," in Proc. of SPIE, 2001, pp. 48-55.

[2] R. Tsai and T. Huang, "Multiframe image restoration and registration," in Advances in Computer Vision and Image Processing: Image Reconstruction from Incomplete Observations, London, 1984, pp. 317-339.

[3] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in Advances in Neural Information Processing Systems, 2006, pp. 291-298.

[4] Q. Yang, R. Yang, J. Davis, and D. Nistér, "Spatial-depth super resolution for range images," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1845-1852.

[5] Y. J. Kil, B. Mederos, and N. Amenta, "Laser scanner super-resolution," in Eurographics Symposium on Point-Based Graphics, 2006, pp. 9-16.

[6] A. N. Rajagopalan, A. Bhavsar, F. Wallhoff, and G. Rigoll, "Resolution enhancement of PMD range maps," Pattern Recogn, vol. 5096, pp. 304-313, 2008.

[7] G. Rosenbush, T. Hong, and R. D. Eastman, "Super-resolution enhancement of flash LADAR range data," in Proc. of SPIE, 2007.

[8] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using Time-Of-Flight depth superresolution," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, pp. 1596-1602.

[9] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "LidarBoost: Depth superresolution for ToF 3D shape scanning," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[10] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," IEEE Trans Image Process, vol. 13, pp. 1327-1344, Oct 2004.

[11] M.-C. Chiang and T. E. Boult, "Efficient image warping and super-resolution," in Proc. of the Third IEEE Workshop on Applications of Computer Vision, 1996, pp. 56-61.

[12] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," IEEE Trans Image Process, vol. 6, pp. 1646-1658, Dec 1997.

[13] S. Baker and T. Kanade, "Super resolution optical flow," The Robotics Inst., Carnegie Mellon Univ. Technical Report CMU-RI-TR-99-36, 1999.

[14] Z. Yuan, P. Yan, and S. Li, "Super resolution based on scale invariant feature transform," in International Conference on Audio, Language and Image Processing, 2008, pp. 1550-1554.

[15] B. K. P. Horn and B. G. Rhunck, "Determining optical-flow," Artif Intell, vol. 17, pp. 185-203, 1981.

[16] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in Proc. of the International Joint Conference on Articial Intelligence, Vancouver, British Columbia, 1981, pp. 674-679.

[17] D. G. Lowe, "Object recognition from local scale-invariant features," in International Conference on Computer Vision (ICCV), Corfu, Greece, 1999, pp. 1150-1157.