

A NEW SUPER RESOLUTION TECHNIQUE FOR RANGE DATA

Valeria Garro, Pietro Zanuttigh, Guido M. Cortelazzo

University of Padova, Italy

ABSTRACT

Current Time-of-Flight matrix sensors allow for the acquisition of range maps, i.e. surface data, at video rate. This possibility is rather intriguing in light of its applications to 3D video and general motion capture. Unfortunately the resolution of the range maps acquired by such devices is quite limited specially if compared with standard color cameras. This makes highly desirable methods for improving the resolution of range maps. The characteristics of range data are completely different from those of images since geometry and light intensity are physically different quantities, therefore methods suited to image interpolation do not perform well with range data. This work presents an original interpolation technique that exploits side information from a standard color camera to increase the resolution of range maps and demonstrates its performance.

Index Terms— Depth map super resolution, range images enhancement

1. INTRODUCTION

In recent years a number of range measuring devices have been developed for 3D data acquisition. There exists a variety of devices with different characteristics that can be chosen according to application needs. For example laser range cameras measure the scene line by line by way of mechanical scanning. Similarly range cameras using light patterns acquire the scene by the projection of a numbers of such patterns. Both types of devices are clearly not suited for the acquisition of dynamic scenes. Range cameras using matrix arrangements of Time of Flight (ToF) sensors can instead obtain 3d measurement of dynamic scenes. These devices employ different technologies. ToF cameras like Mesa Imaging's Swiss Ranger [1] and Canesta [2] compute the points' depth by emitting light signals and measuring the time delay the reflected light signals takes to reach the sensor again. Other range cameras like Zcam [3] implement a method based on optical shutter technology.

A major advantage of matrix ToF sensors is that they can acquire untextured geometry where passive methods (stereo) fail. Probably their main current drawback is their limited resolution (e.g. Swiss Ranger works with a 176×144 pixel array, Canesta sensor's resolution is 160×120). To overcome

this issue this paper presents a novel super resolution method for generic range maps.

Previous works exploited additional information to improve depth map's resolution combining ToF sensor with one or two high-resolution RGB cameras; the main idea is that depth discontinuities in range map are often related to color and brightness changes in the corresponding color image.

Some of these techniques use a probabilistic approach: in [4] the enhancement is based on Markov Random Field (MRF) formulation of depth data solved via conjugate gradient; another approach [5] models a cost volume of depth probability and iteratively applies bilateral filtering providing a spatial resolution enhancement of $100\times$ ($10\times$ width and $10\times$ height).

Another recent method [6] uses exclusively depth maps, without color image aid: many low-resolution depth maps of the same scene are aligned and merged together in order to obtain a single depth map with improved resolution. This method is restricted to static scenes' acquisition.

We present an alternate method for range maps super resolution not based on a probabilistic formulation: our method exploits additional information given by a single RGB camera registered with the range camera. The key feature is the estimation of surface's patches with a novel interpolation strategy explicitly targeted to depth data. An ancillary image represents information which can be collected very easily and at low cost.

The structure of this paper is the following: Section 2 gives an overview of our method and the details of our interpolation algorithm, Section 3 discusses the experimental results, and Section 4 draws the conclusions.

2. PROPOSED SUPER RESOLUTION TECHNIQUE

We propose an alternate super resolution method for range maps which works with a single additional registered RGB camera. Our method relies on assumptions similar to those of [4] and [5]: depth map discontinuities and color or brightness changes in the related color image are correlated, especially in presence of edges. As previously said, our problem's formulation differs from other methods because it does not use a probabilistic framework, instead it extends the depth estimation procedure developed for the compression scheme of [7].

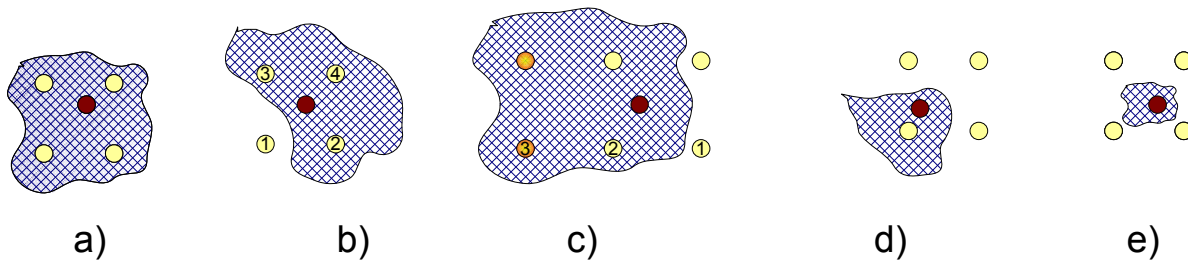


Fig. 1. Grid samples: the 5 possible configurations

The proposed super resolution method encompasses various steps. In the first step the color image is pre-processed with an efficient graph-based segmentation algorithm [8] in order to identify the main surfaces of the scene. The connected components created in the output image, defined as regions with uniform pixel values, correspond to surface regions that ideally belong to the same scene object. These regions can be approximated by a set of planar patches.

The second step of the proposed super resolution scheme is the projection of the points of the low-resolution depth map on the segmented color image. With this procedure a low-resolution depth grid is laid upon the color image. The elements of this grid are the depth values of the initial low-resolution depth map.

The proposed approach is based on the assumption that for each segmented region a set of projected depth points is available. For this reason one must prevent the creation of too small segmented regions by the segmentation algorithm. The heart of the proposed method is the interpolation step described in detail in Section 2.1 and 2.2

2.1. Interpolation algorithm: basic structure

After the above mentioned pre-processing steps each pixel of the segmented color image can therefore be surrounded by up to 4 samples of the low-resolution grid belonging to the same region. The value of each pixel of the high-resolution depth map can be estimated by interpolating the values of the grid samples that belong to the same region.

There are 5 possible cases (excluding the trivial case of samples that correspond to grid points), exemplified in Fig.1 (other symmetrical or rotated configurations are possible). The interpolation algorithm works as follows according to the different pixel configurations:

- a) (**Fig.1a**) All the four samples surrounding the pixel to be estimated (shown in red) are inside the same region. In this case the pixel value is obtained from the value of the 4 closest grid samples by a modified bilinear interpolation. This approach is consistent with the assumption that most surface patches are planar. Experi-

mental proofs with splines and other cubic interpolation schemes showed that in most practical cases these techniques do not provide any improvement on the surface reconstruction accuracy. The details of the adopted interpolation scheme are explained in Subsection 2.2.

- b) (**Fig.1b**) The pixel to be estimated is surrounded only by three samples belonging to the same region the fourth grid sample is out of the region. In this case we compute the value that the fourth grid sample would have if it was in the region by assuming that it is on the plane defined by the other three points. Referring to the pixel position numbering of Fig.1b, this corresponds to estimating the value of \hat{g}_1 as

$$\hat{g}_1 = g_2 + g_3 - g_4 \quad (1)$$

Then we calculate the pixel value using modified bilinear interpolation (with the 3 samples in the region and the fourth point \hat{g}_1 previously computed) as in the first case.

- c) (**Fig.1c**) The pixel to be estimated has two neighbors in the region and two outside (this typically happens when it is close to an edge). The two missing points are computed by assuming that each of them lies on the line passing through the closest available point and the symmetrical point with respect to the available one (shown in orange). With the notation of Fig.1c, for example, we obtain

$$\hat{g}_1 = 2g_2 - g_3 \quad (2)$$

If the symmetrical point was also outside the region the “line estimation” is not valid and the closest point value would be chosen as reasonable estimate of g_1 .

As in the previous case once the two missing grid samples are obtained the pixel depth value to be estimated is computed by modified bilinear interpolation.

- d) (**Fig.1d**) The pixel to be estimated has just one neighbor grid sample in the same region: the value of this sample is taken as its estimate.

e) (**Fig.1e**) The pixel to be estimated is inside a very small region and none of the 4 closest grid samples is inside it (quite uncommon but possible, it can be prevented with a proper choice of the segmentation's parameters). In this case the value of the nearest sample is taken as its estimate.

An important aspect of the proposed procedure is that the high-resolution range map values are always estimated from grid samples in the same segmented region of the range map's pixel to be evaluated, while grid samples outside the region are never used (excluding the case where no sample is in the region). This ensures that steep edges do not get blurred unless they are not detected by the segmentation.

2.2. Provision for handling unreliable segmentation information

In order to compute the depth value g of a new pixel of the high-resolution depth map, steps a), b) and c) of the interpolation scheme of Section 2.1 use bilinear interpolation of the four closest grid's values $g_i \mid i = \{1, \dots, 4\}$ surrounding it (see Fig.2), as

$$\hat{g} = \frac{(h_x f_y)g_1 + (f_x f_y)g_2 + (h_x h_y)g_3 + (f_x h_y)g_4}{(x_b - x_a)(y_b - y_a)} \quad (3)$$

where $f_x = x - x_a$, $f_y = y_b - y$, $h_x = x_b - x$ and $h_y = y - y_a$ (see Fig.3).

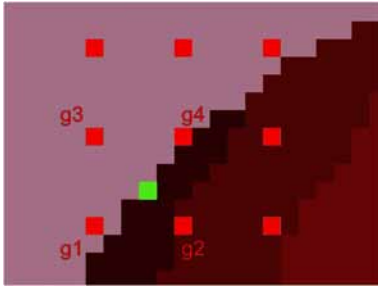


Fig. 2. Example of depth grid configuration: the green pixel is surrounded by the four red grid's point g_1, g_2, g_3, g_4

This approach would work quite well if each depth sample was assigned to the correct region. In practice some samples can be assigned to a wrong region due to the limited precision of the calibration of the acquisition setup or due to the segmentation errors. Such misassignments typically concern pixels in proximity of edges. Indeed it may happen that a pixel of the low-resolution depth map close to an edge may be projected on the wrong segmented region as shown in Fig.4. In order to overcome the problems caused by the pixels of the low-resolution depth map assigned to the wrong region we devise a way of underweighting their contribution in (3). To this purpose two parameters ρ and Φ are introduced.

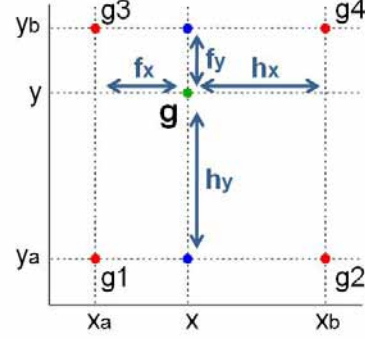


Fig. 3. Bilinear interpolation configuration

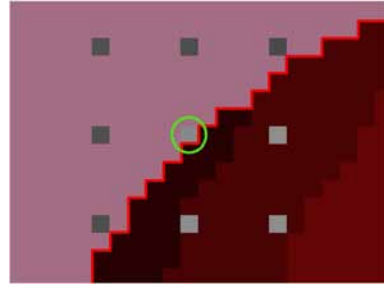


Fig. 4. Example of invalid depth value attribution: the green circled depth value is projected on the background pink region but its depth value corresponds to a foreground object.

Parameter ρ_{g_i} applies to each pixel g_i of the original low-resolution range map and essentially evaluates its trustworthiness upon its distance from the edges of the segmented regions, obtained from the color image. In particular it is defined as:

$$\rho_{g_i} = \frac{\min \{d_{g_i}, gsize\}}{dmax} \quad (4)$$

where:

- d_{g_i} is the distance of g_i to the closest edge,
- $gsize$ is the ratio between color image resolution and initial depth map resolution,
- $dmax = \max_{i=1, \dots, 4} \{ \min_{i=1, \dots, 4} \{d_{g_i}, gsize\} \}$.

The idea is that samples closer to the edge should receive a lower weight to account for the fact that they are less reliable. For every four-samples configuration of the low-resolution range map it is useful to identify the furthest from the edge. Such a pixel is denoted as g_M where index M is defined as the argument of $dmax$, i.e.,

$$M = \arg \left\{ \max_{i=1, \dots, 4} \left\{ \min_{i=1, \dots, 4} \{d_{g_i}, gsize\} \right\} \right\}. \quad (5)$$

Pixel g_M is instrumental for the definition of the second parameter Φ which considers the reliability of the grid values \hat{g}_i estimated by equations (1) and (2). A threshold function is used in order to give them a lower weight if the difference $\delta = |g_M - \hat{g}_i|$ between the grid point g_M and the estimated grid point value \hat{g}_i is greater than a threshold t_1 ; such a weight decreases linearly and it is set to 0 if difference δ is greater than a second threshold t_2 (see Fig.5). Namely for each pixel of the original low-resolution range map parameter $\Phi_{g_i} : [0, 1] \rightarrow [0, 1]$ is defined as:

$$\Phi_{g_i}(\delta) = \begin{cases} 1 & \text{if } \delta < t_1, \\ -\left(\frac{1}{t_2-t_1}\right)\delta + \left(1 + \frac{t_1}{t_2-t_1}\right) & \text{if } t_1 \leq \delta \leq t_2, \\ 0 & \text{if } \delta > t_2. \end{cases} \quad (6)$$

Thresholds $t_1, t_2 \in [0, 1]$ are set to suitable values experimentally derived.

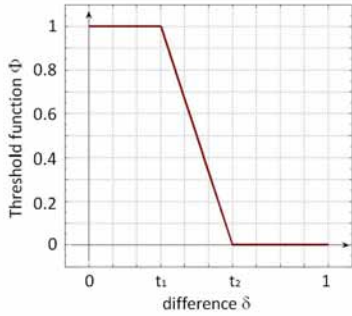


Fig. 5. Threshold Function Φ .

Bilinear interpolation (3) in the proposed super resolution algorithm is replaced by:

$$\hat{g} = \frac{(\rho_{g_1} \Phi_{g_1})(h_x f_y) g_1 + (\rho_{g_2} \Phi_{g_2})(f_x f_y) g_2}{\nu} + \frac{(\rho_{g_3} \Phi_{g_3})(h_x h_y) g_3 + (\rho_{g_4} \Phi_{g_4})(f_x h_y) g_4}{\nu} \quad (7)$$

where:

$$\nu = (\rho_{g_1} \Phi_{g_1})(h_x f_y) + (\rho_{g_2} \Phi_{g_2})(f_x f_y) + (\rho_{g_3} \Phi_{g_3})(h_x h_y) + (\rho_{g_4} \Phi_{g_4})(f_x h_y)$$

3. EXPERIMENTAL RESULTS

The proposed method was tested on several scenes. It is worth noting that parameter $gsize = \frac{\text{image-size}}{\text{depth-map-size}}$ essentially sets the interpolation factor in one dimension. Parameters t_1 and t_2 of threshold function Φ were set to 0.1 and 0.4 respectively.

For validation purposes our experimentation began with synthetic data, so that the output of the proposed method could be compared against a ground truth. The proposed algorithm was tested in order to increase the resolution of an original

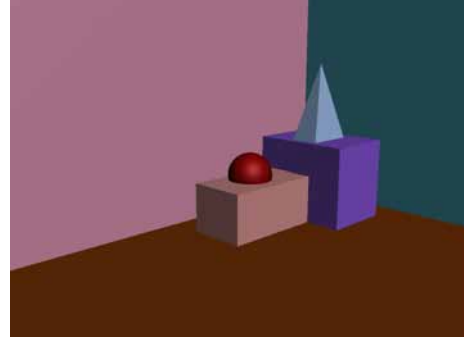


Fig. 6. Color rendering of the synthetic scene

Table 1. Experimental configurations with initial depth map of 128×96 pixels, MSE measurements with corresponding ground truths

	image resolution	depth map resolution	$gsize$ range map interp. factor	MSE
case 1	1024×768	128×96	8	3.66
case 2	1280×960	128×96	10	3.04
case 3	1536×1152	128×96	12	2.61
case 4	1920×1440	128×96	15	2.26
case 5	2304×1728	128×96	18	1.88
case 6	2560×1920	128×96	20	1.83

128×96 range map of the synthetic 3D scene of Fig.6 to six higher resolutions associated to different color images. Table1 reports the mean square error (MSE) measurements between our output and the corresponding ground truth. By comparing the six different cases we notice a steady improvement with resolution's increase. This is due to the fact that the average reliability of the grid points is proportional to the range map interpolation factor and to the resolution of the ancillary image used for segmenting the scene's regions. This fact does not come as a surprise since the higher is the image resolution and the range map interpolation factor the better are the segmented regions with respect to the information associated to the original low-resolution range image. In this connection it is worth pointing out that range map interpolation factor and image resolution are not independent and should be chosen accordingly for maximum performance.

A visual comparison of the initial depth map (128×96) and the interpolated output depth map of case 1 (1024×768) is provided in Fig.7. In Fig.8 some details of the output depth map of case 6 are compared against the low-resolution depth map.

The proposed method with the synthetic data of Fig.6 was compared against standard image interpolation methods: linear interpolation, cubic interpolation and Lanczos interpolation. Table2 and Fig.9 show the performances.

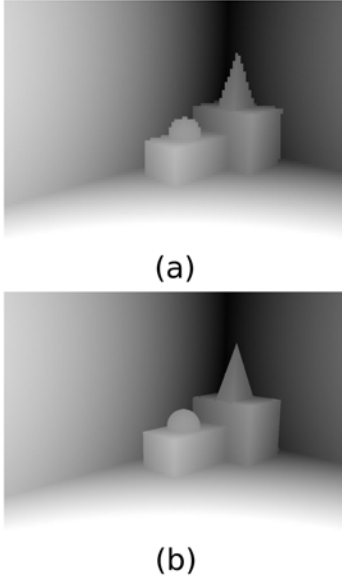


Fig. 7. (a) Initial depth map 128×96 ; (b) output depth map 1024×768 .

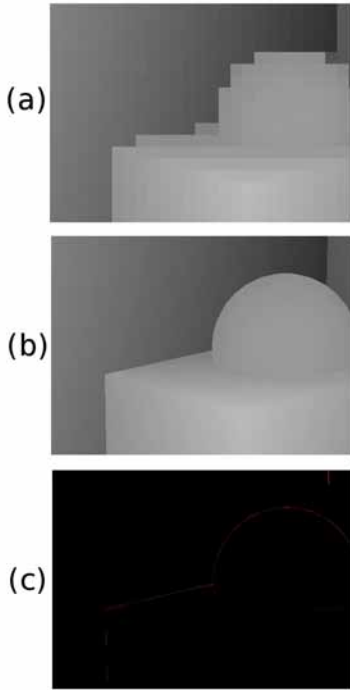


Fig. 8. Sphere and corner detail - (a) initial depth map 128×96 ; (b) output depth map 2560×1920 ($gsize = 20$); (c) difference with respect to the ground truth.

The proposed method “ah hoc” for range data is clearly superior to standard image interpolation techniques. The MSE of the proposed approached in the worst case is 3.66 where other methods never go below 19.

Table 2. Experimental results: Mean Square Error comparison.

Output Resolution	Linear Interp.	Cubic Interp.	Lanczos Interp.	Proposed Method
1024×768	20.81	22.59	21.53	3.66
1280×960	20.50	22.14	21.16	3.04
1536×1152	20.00	21.44	20.62	2.61
1920×1440	19.59	21.17	20.07	2.26
2304×1728	19.70	21.31	20.31	1.88
2560×1920	19.64	21.19	20.21	1.83

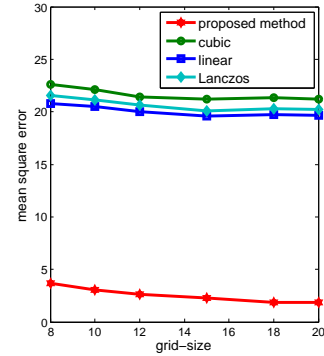


Fig. 9. Mean Square Error comparison.

The proposed modified bilinear interpolation prevents gradual shading’s effects typical of unreliable grid values in the estimate of the surface’s patches, Fig.10 and Fig.11 compare the use of standard bilinear interpolation (3) against the proposed expression (7) in steps a) b) c) of Section 2.1 on some details of the range map of Fig.7b. The artifacts introduced by segmentation and calibration issues are clearly shown by Fig.10a, Fig.11a, Fig.11c. They do not occur if in the interpolation steps standard bilinear interpolation is replaced by (7) as shown by Fig.10b, Fig.11b and Fig.11d.

The proposed method was also tested on real data con-

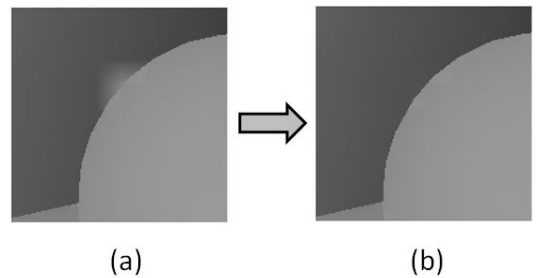


Fig. 10. Comparison between our method and simple bilinear interpolation - a) bilinear interpolation, b) proposed method.

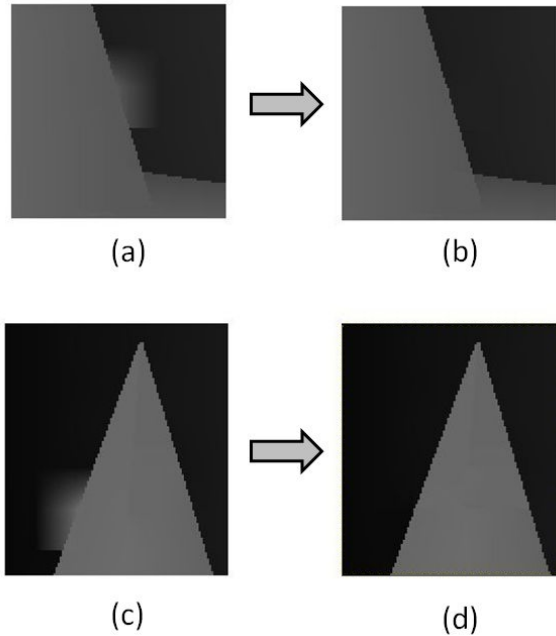


Fig. 11. Comparison between our method and simple bilinear interpolation - a) c) bilinear interpolation, b) d) proposed method.

cerning a scene with a human actor (see Fig.12). The initial depth map was obtained with passive methods and it has a resolution of 128×96 . The 3D geometry of the model was recovered using the method presented in [9] that is based on the combined use of stereo and shape from silhouette information. The color image of the scene has a resolution of 1024×768 pixels ($gs_{size} = 8$).



Fig. 12. Real Scene - (a) color image; (b) segmented image.

As we do not have available ground truth information of this depth map we can not make a quantitative evaluation with mean square error measurements but we can only rely on visual comparisons. Fig.15 and Fig.16 compare the high-resolution output depth map obtained by the proposed method against standard methods of image upscaling. Again the performance of the proposed method is clearly superior.

Computation is very efficient due to the fact that each step involves simple low computation operations. As expected,



Fig. 13. Depth maps' comparison - (a) initial depth map; (b) output depth map.

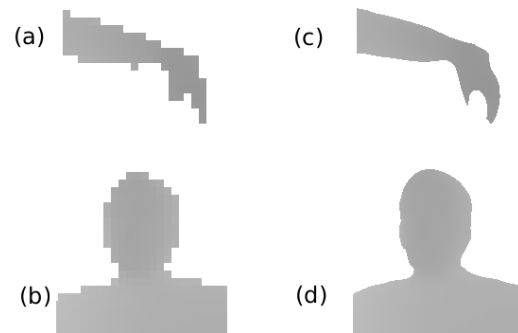


Fig. 14. Depth maps' comparison - details - (a) - (b) arm and head details of initial depth map ; (c) - (d) arm and head details of output depth map.

such a cost increases with the use of higher resolution color images.

4. CONCLUSIONS

The interest in super resolution techniques targeted to range maps is fostered by recent matrix ToF devices capable to capture surface data at video rate. Unfortunately image interpolation methods are not suited to this application. In the literature some methods have been proposed based on probabilistic models of range data [4] [5]. Current video cameras' costs are quite lower than those of matrix ToF devices while their resolution is quite higher, therefore it seems rather sensible with respect to costs/benefits trade-off using imagery in order to assist range data interpolation.

This work proposes an original technique for increasing the resolution of range maps based on the information of a color image registered on the depth map. Since interpolation artifacts typically occur at geometry edges the basic idea behind the proposed technique is to use the color image in order to locate the geometrical edges of the scene. Interesting enough since the resolution of the color image is rather higher than that of the original low-resolution depth map the edge

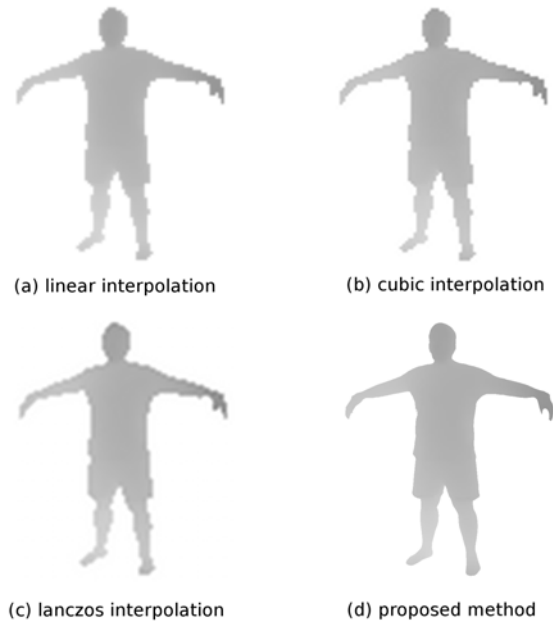


Fig. 15. Visual comparison of the whole body.

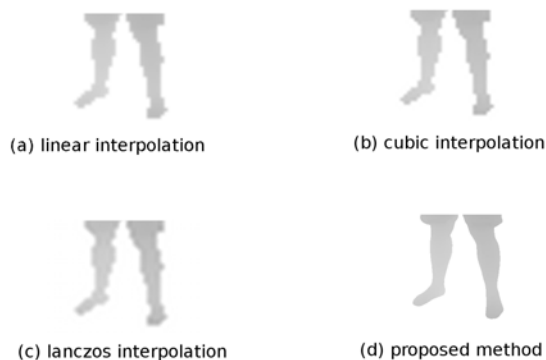


Fig. 16. Visual comparison of legs' detail.

information provided by the color image is rather reliable. Experimental data show that the proposed technique outperforms standard image interpolation methods used on range data. The proposed techniques versus super resolution methods for range data based on probabilistic approaches has the advantage that it is computationally very efficient. In principle the performance of the proposed techniques is conditioned by the precision of the registration between the range map and the color image and by the performance of the standard segmentation step, however the techniques developed in Section 2.2 make the proposed interpolation scheme more robust to this issues. Segmentation methods based on graph-cuts experimentally appear rather adequate for the considered tasks. However, in presence of highly textured surfaces, segmentation methods based on the combined use of depth and texture

information [10, 11] can be adopted in order to avoid an over-segmentation effect.

Current work aims at making the proposed procedure as robust as possible against misregistration and segmentation faults. Another research direction examines the advantages of using a pair of images instead of a single one as an aid for range map super resolution: indeed the system complexity cost would increase only moderately but the image pair would make available also geometric information which could not only assist super resolution but also directly integrate the data provided by the matrix ToF sensors in low reflective areas or noisy operation conditions.

5. REFERENCES

- [1] "Mesa imaging swiss ranger sr4000," <http://www.mesa-imaging.ch>.
- [2] "Canesta inc, canestavision electronic perception development kit," <http://www.canesta.com>.
- [3] "3dv systems zcam," <http://www.3dvsystems.com>.
- [4] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*, Cambridge, MA, 2005, MIT Press.
- [5] Q.X. Yang, R.G. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," in *Proceedings of CVPR07, 2007*, pp. 1–8.
- [6] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "High-quality scanning using time-of-flight depth superresolution," in *Proceedings of CVPRW08, 2008*, pp. 1–7.
- [7] P. Zanuttigh and G.M. Cortelazzo, "Compression of depth information for 3d rendering," in *Proceedings of 3DTV, 2009*.
- [8] Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [9] Luca Ballan, Nicola Brusco, and Guido Maria Cortelazzo, "3D passive shape recovery from texture and silhouette information," in *2nd IEE European Conference on Visual Media Production, CVMP*, London, UK, November 2005.
- [10] Ryan Crabb, Colin Tracey, Akshaya Puranik, and James Davis, "Real-time foreground segmentation via range and color imaging," *Computer Vision and Pattern Recognition Workshop*, vol. 0, pp. 1–5, 2008.
- [11] Y. Ma, Stewart Worrall, and Ahmet M. Kondoz, "Automatic video object segmentation using depth information and an active contour model," in *Proceedings of MMSP, 2008*, pp. 910–914.