On the asymmetric view+depth 3D scene representation

ON THE ASYMMETRIC VIEW+DEPTH 3D SCENE REPRESENTATION

Mihail Georgiev, Atanas Gotchev

Department of Signal Processing, Tampere University of Technology, Finland

ABSTRACT

In this work we promote the asymmetric view + depth representation as an efficient representation of 3D visual scenes. Recently, it has been proposed in the context of aligned view and depth images and specifically for depth compression. The representation employs two techniques for image analysis and filtering. A super-pixel segmentation of the color image is used to sparsify the depth map in spatial domain and a regularizing spatially adaptive filter is used to reconstruct it back to the input resolution. The relationship between the color and depth images established through these two procedures leads to substantial reduction of the required depth data. In this work we modify the approach for representing 3D scenes, captured by RGB-Z capture setup formed by non-confocal RGB and range sensors with different spatial resolutions. We specifically quantify its performance for the case of low-resolution range sensor working in low-sensing mode that generates images impaired by rather extreme noise. We demonstrate its superiority against other upsampling methods in how it copes with the noise and reconstructs a depth map with good quality out of very low-resolution input range image.

1. INTRODUCTION

'View plus depth' is a 3D scene representation, which combines the photographic quality of color images with information about the scene geometry encoded by distance (depth) maps [1]. It is used in a number of applications such as 3D video coding, free viewpoint rendering, and mixing real with synthetic scenes (augmented reality), to name a few. In 3D video coding, the representation is instrumental for decoupling the capture format from the display format and for serving various stereoscopic and multiview displays. In free-viewpoint view rendering, it helps in providing near-continuous parallax, while in augmented reality it assists the correct insertion of synthetic objects within the real scene or vice versa.

The depth modality is provided by either passive sensing techniques, employing depth from stereo estimation [2], or by active sensing, allocating dedicated range (distance) measurement devices. The latter utilize techniques such as structure light [3] or time-of-flight near-infrared continuous wave beamers and sensors [4]. In most cases the targeted output format is V+D, where the depth is aligned to the color image and the two modalities have the same spatial resolution [1]. In the format, the two modalities are essentially different. While the color modality represents the color texture variations of the objects, the depth modality is a piecewise smooth function which represents the gradual change of distances between the camera and scene objects. However, it also has edges well aligned with the edges in the color image as objects being at different depths create those. The piecewise-smooth behavior of the depth has been taken into account when designing depth compression methods such as methods based on platelets [5], anisotropic threes [6], constant-value segments [1], or contour encoding [7]. The correlation between the depth and color images has been exploited by related prediction structures [8], e.g. edges from the color image estimate the edges in the depth image, or by designing dictionaries for joint intensity-depth sparsification [9].

Recently, we have proposed an asymmetric V+D representation, which substantially reduces the amount of data needed to represent the depth modality [10]. The depth is decimated based on an optimal super-pixel segmentation of the color image to a regular structure, which resembles a low-resolution depth image. The optimization of the super-pixel map is done with respect to a bilateral depth reconstruction filter, which reconstructs the decimated depth map back to its original resolution. We demonstrated the feasibility of the approach for compressing aligned V+D images with the same resolution [10]. In this work, we study the performance of that model for the case when a ToF sensor provides the depth map. The difference with the previous setting is that the sensed data appears noisy and on a different image plane. We modify the method accordingly and perform experiments to quantify how much the measurement noise invalidates the representation model and whether the model can cope with certain amount of noise. The paper is organized as follows. The proposed asymmetric V+D representation is presented in Section 2. Section 3 discusses the peculiarities of RGB+ToF sensing setup and the modifications to the representation it induces. Experimental results and related discussion are presented in Section 4, followed by conclusions in Section 5.

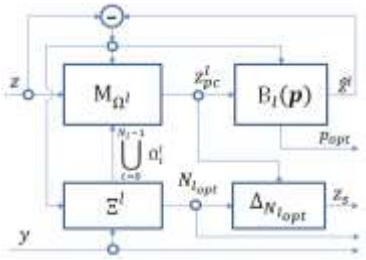

Fig 1. Flow chart of proposed depth sparsification scheme

2. ASSYMETRIC V+D REPRESENTATION

2.1. Overview of approach

Consider a view plus depth 3D scene representation where the color view modality is represented by a color image in e.g. RGB color space $\boldsymbol{y}(\boldsymbol{n}) = [y^R(\boldsymbol{n}), y^G(\boldsymbol{n}), y^B(\boldsymbol{n})]$ and the associated depth modality is represented by a gray scale image $z(\boldsymbol{n})$, where $\boldsymbol{n} = [n_1, n_2]$ is a spatial variable, $\boldsymbol{n} \in \mathrm{N}$, N being the image domain. We aim at finding a sparse representation of the depth, related with the aligned color image. The sparsification is sought in spatial domain. The goal is to output a low-resolution regular (isotropic) depth image z_s, which structure is driven by corresponding color image segmentation in terms of super-pixels. The flowchart of the system, implementing our approach is given in Fig. 1. At each iteration l, the color image is segmented by a super-pixel segmenting operator Ξ^l. A masking operator M_{Ω^l} replaces depth values at each segment Ω_i^l by a constant thus creating a piecewise-constant depth image z_{pc}^l. A reconstruction filter B_l uses the latter to reconstruct an estimate $\hat{z}$ of the original depth. The difference $z - \hat{z}$ controls the optimization of the filter parameters, the choice of constant depth values in super-pixel segments, and the next segmentation iteration. Upon reaching a sparsification criterion (or a bit budget limit), the optimal segmentation is selected and the irregularly segmented depth is turned to a regular low-resolution depth image z_s by the corresponding operator $\Delta_{N_{l_{opt}}}$. The main blocks of the scheme are described in more details in the following subsections.

2.2. Color image segmentation in terms of super-pixels

By super-pixels we denote isotropic image segments having compact representation. They represent homogenous areas in terms of color and texture and behave as raster pixels on a low-resolution near-regular grid. Super-pixels can be generated by various segmentation approaches. We refer to two such approaches, namely: *Simple Linear Iterative Clustering* (SLIC) [11], and *Super-pixels Extracted via Energy-Driven Sampling* (SEEDS) [12]. Examples of super-pixel segmentation are given in Figure 2. The structure of the super-pixel image segmentation is determined by the input (desired) number of segments. Given that number, the structure is fully reproducible in terms of segment areas and indexing [13].

With reference to Fig. 1, consider several successive color image segmentation stages $l = 0, 1, \ldots L-1$ resulting in increasing (refined) number of super-pixels $N_0 < N_1 < .. N_l < \ldots$ At stage l, the segmenting operator Ξ^l outputs N_l super-pixels indexed by $i, i = 0, 1, \ldots N_l - 1$, where the corresponding segments are denoted by $\Omega_i^l, \bigcup_{i=0}^{N_l-1} \Omega_i^l = \mathrm{N}$.

In our approach, super-pixels are the blocks, which relate the color image with the associated depth map. As super-pixels are delineated along edge shapes between color textures and objects [13], it is expected that the same edge shapes between objects be presented in the depth map.

2.3. Piecewise-constant depth modeling and reconstruction

Under the assumption for a piecewise-constant depth within each super-pixel segment, the depth values within the segment are replaced by a constant value $z_{pc}(\boldsymbol{n}) = z_i, \boldsymbol{n} \in \Omega_i^l$. The constant values are selected in such a way as to minimize the mean squared error (MSE) between the original depth and the depth reconstructed by the filter B_l. The latter should impose locally-adaptive regularization in order to reflect the spatial relation between the color and depth images. Advanced filters, both local and non-local can be adapted for this purpose [14, 15, 16]. In our experiments, we use the cross-bilateral filter [17] in the following setting

$$z_{i,opt} = \operatorname{argmin} \left(\sum_{\boldsymbol{n} \in \Omega_i^l} z(\boldsymbol{n}) - \hat{z}(\boldsymbol{n}) \right)^2, \text{ where}$$

$$\hat{z}(\boldsymbol{n}) = \frac{\sum_{\boldsymbol{m} \in \mathrm{R}} e^{\frac{\|\boldsymbol{n}-\boldsymbol{m}\|}{\alpha}} e^{\frac{\|\boldsymbol{y}(\boldsymbol{n})-\boldsymbol{y}(\boldsymbol{m})\|}{\beta}} z_{pc}(\boldsymbol{n})}{\sum_{\boldsymbol{m} \in \mathrm{R}} e^{\frac{\|\boldsymbol{n}-\boldsymbol{m}\|}{\alpha}} e^{\frac{\|\boldsymbol{y}(\boldsymbol{n})-\boldsymbol{y}(\boldsymbol{m})\|}{\beta}}}. \quad (1)$$

For each segmentation stage, the parameters of the filter $\boldsymbol{p} = \{\alpha, \beta\}$ and the filter size R are also optimized to give a minimum MSE [18].

2.4. Regular low-resolution depth image

Upon reaching a desired quality of the reconstructed depth or a targeted bit-budget, an optimal number of super-pixels $N_{l_{opt}}$ at segmentation stage l_{opt} is selected. The corresponding piecewise-constant segmented depth is converted to a regular image z_s. Each pixel of this image corresponds to a rectangular tile, approximately covering one super-pixel of constant depth and takes its value. The size of the rectangular tile is selected to have the best aspect ratio determined by the respective irregular super-pixel and to cover relatively the same area.

The above proposed asymmetric view + depth 3D representation has the following advantages:

- The sparsified depth map is a regular image, which is easy to interpret and compress. It resembles the output of sensed depth maps;
- In the representation, there is no need to use learned dictionaries of depth representation atoms or other sparsifying transforms – everything comes from the two computing modules: the super-pixel segmentation and the spatially-adaptive reconstruction filter;
- This combination establishes the structural relationship between the view and the depth. The original-resolution depth can be reconstructed by a regularized filtering, involving the color modality.
- The main benefit of super-pixels in contrast to other segmentation approaches is that the partitioning information (e.g. contours, edges, coordinates, indexing) is not required since the segmentation is fully reproducible.

An example is given in Fig. 2. A color frame from the *Breakdancers* dataset is segmented in varying number of super-pixels, the corresponding depth maps are downsampled and converted into regular maps and then reconstructed. Resulting PSNRs are given on top of the reconstructed depth images.

3. 3D SCENE SENSED BY RGB + TOF CAMERAS

3.1. Principle of operation of ToF range sensor

In active 3D scene sensing, an RGB sensor provides the color modality, while an active range sensor provides the depth modality. It this work we specifically consider the depth as provided by a Time-of-flight (ToF) sensor. A typical ToF device consists of a beamer, an electronic light modulator and a sensor chip. An array of light-emitting diodes (LED) operating in near-infrared wavelengths (*e.g. 850 nm*) forms the beamer. It radiates a continuously-modulated harmonic signal which illuminates the scene. Sensor elements (pixels) sense the light reflected from object surfaces by collecting charges during a selected integration time I_T (e.g. 2 ms). For each pixel, the range data is estimated in relation to the phase-delay between the emitted and sensed (reflected) signals [4]. The phase-delay estimate is obtained as a discrete cross-correlation of successive measurements taken at equal intervals during the same modulation period. Mixed signal components: amplitude A and phase φ are estimated from the raw measurements [4]. The sensed distance d is proportional to the phase, while the variance of distance measurements σ_d^2 is proportional to the square inverse of the amplitude and the amplitude is inversely proportional to a monomial of distance [19]:

$$d \propto \varphi,\ \sigma_d^2 \propto {}^{1}/_{A^2},\ A \propto {}^{1}/_{d^\eta},\ \sigma_d \propto d^2\ . \qquad (2)$$

The value of d is calculated after precise calibration of the sensor. The sensed distance signal is prone to measurement errors, caused by low- or multi-reflectance surfaces or ambient light. Furthermore, the ToF sensor might be set to work in so-called low-sensing mode, in an attempt to reduce the power consumption and the sensor's physical dimensions. Specific denoising methods have been developed to cope with this type of noise [20,21,22].

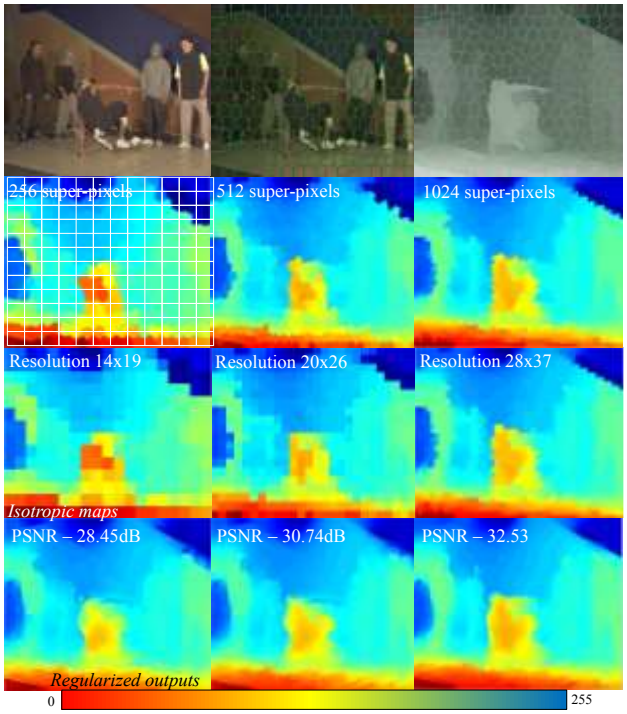

Fig. 2. Depth sparsification by color image segmentation in super-pixels (by columns from top): color image, the same segmented in super-pixels, the depth image with superimposed super-pixel regions; Downsampled depth maps for varying number of super-pixels; Corresponding isotropic maps; Regularized reconstructed output and corresponding PSNRs.

3.2. Sensed asymmetric view + depth

The asymmetric V+D representation discussed in Section 2 assumed that the color and depth images are aligned i.e. they 'see' the scene from the same viewpoint. In the case of active 3D scene sensing, however, the RGB and depth sensors are non-confocal. They sense the scene from two slightly different perspectives and with different field of view. There is a projection relation mapping the coordinates of the ToF sensor on the RGB sensor plane.

The spatial resolution of the sensors is also different. While the conventional color (RGB) cameras are with high resolution, the ToF sensors are with rather low resolution. Their sensor elements have larger surface compared to RGB elements (e.g. 150 µm vs. 8 µm) in order to cope with the sensitivity in the near-infrared light wavelength region. The geometric relations between the sensors can be formalized as follows: Consider the range map pixel $d(m_1, m_2)$. The homogeneous world coordinates of the point are given by

$$Z = f_{ToF}\, d(m_1, m_2) \Big/ \sqrt[2]{f_{ToF}^2 + m_1^2 + m_2^2}$$
$$U = m_1 Z / f_{ToF}, \qquad V = m_2 Z / f_{ToF}, \qquad w = 1, \tag{3}$$

where f_{ToF} is the focal length of the range sensor. The point is then reprojected on the RGB sensor plane

$$(u, v, w) = \boldsymbol{KP}[U\ V\ Z\ w]^T$$
$$(n_1', n_2') = (u, v)/w\,, \tag{4}$$

where $\boldsymbol{P}$ is the projection matrix relating the optical centers of the two sensors (camera extrinsics) and $\boldsymbol{K}$ is the matrix with the intrinsic parameters of the RGB sensor. Note that a precise joint calibration between the two sensors is needed in order to identify the projection matrix $\boldsymbol{P}$. Discussing such calibration falls outside the scope of this paper.

The projected depth $z_{proj}(\boldsymbol{n}')$, $\boldsymbol{n}' = [n_1', n_2']$ appears at non-uniform coordinate positions with respect to the coordinate grid of the given RGB image $\boldsymbol{y}(\boldsymbol{n})$, $\boldsymbol{n} = [n_1, n_2]$, $\boldsymbol{n} \in \mathrm{N}$. Our proposal is to convert this non-uniform depth map into a regular depth map structurally related with the color image through super-pixel segmentation. An optimal color image segmentation $\Xi: \{\Omega_i\}$, $\bigcup_{i=0}^{N_{opt}-1} \Omega_i = \mathrm{N}$ is found in such a way so that each segment contains at least one point from the projected depth, i.e. $N_{opt} \leq M$, where M is the number of projected depth samples. The projected depth z_{proj} is replaced by one sample per segment $z_\Omega(i) = \max_{\boldsymbol{n}'} \left(z_{proj}(\boldsymbol{n}') \right), \boldsymbol{n}' \in \Omega_i$. Subsequently, the non-uniform depth z_Ω is converted to a depth image being piecewise-constant over the super-pixel regions then to a regular low-resolution depth image con-focal with color one: $z_s = \Delta_{N_{opt}} \mathrm{M}_\Xi z_\Omega$.

4. EXPERIMENTAL RESULTS AND DISCUSSION

We characterize the performance of the so-proposed asymmetric V+D representation by experiments on a synthetic photorealistic scene with known ground true depth. A non-confocal capture system is simulated. The color and ToF cameras have the same field of view of 50° and are positioned at 6 cm baseline. The spatial resolution of the RGB camera is set to 1024x768 pixels and the depth range is set to 0÷7.5m. The size of the range sensor pixel and the corresponding resolution vary as given in Table 1. The scene is shown in Fig. 3. The results are compared in terms of *PSNR* between reconstructed and ground true depth at the resolution of the color camera. Occluded pixels have been excluded from the evaluations.

In the first experiment we quantify the performance of the super-pixel segmentation found with respect to the sensed reprojected depth in noise-free conditions. The segmentation determines the nearest-neighbor

Table 1. Varying range sensor resolutions, pixel sizes, and scale factors wrt color image resolution

Resolution	640x480	320x240	160x120	80x60	40x30
Pixel size [μm]	50	100	200	400	800
Scale factor	2.56	10.24	~96	~163	~655

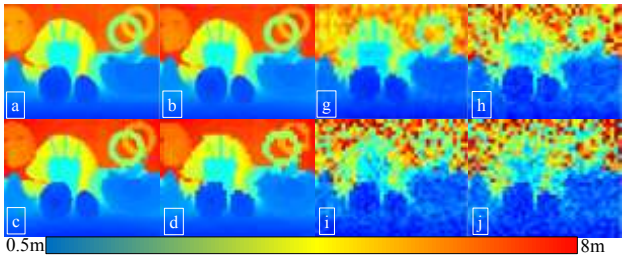

Fig 3. Rendered output of low-resolution range sensor for spatial resolutions of: a) 640x480, b) 160x120, c) 80x60, d) 40x30 pixels; simulated ToF noise added to the 40x30 sensor output for σ_D^2: g) 0.75 m, h) 2 m, i) 2.5 m, and j) 3 m.

interpolation from non-uniform depth to piecewise-constant depth to be used in the depth map regularized reconstruction. We apply an iterative reconstruction of Richardson type:

$$\hat{z}^{l+1} = \mathrm{B}\left(\hat{z}^l + \lambda T_\mathrm{A}\left(z_{proj} - L\hat{z}^l \right) \right), \quad \mathrm{A} = \{\Omega, \mathrm{V}, \Delta\}. \tag{5}$$

In the equation, L denotes a piecewise-linear interpolator, which interpolates depth values at the non-uniform grid points $\boldsymbol{n}'$ given the depth at the uniform grid $\boldsymbol{n}$; T_A is some upsampling operator, which upsamples the depth from the grid points $\boldsymbol{n}'$ to the grid points $\boldsymbol{n}$; B is a regularized reconstruction filter, e.g. a bilateral filter, and λ is a relaxation parameter. Two alternatives to the super-pixel based upsamping T_Ω are considered. T_V utilizes Voronoi cells V_i, fitted around each non-uniform coordinate $\boldsymbol{n}_i'$ [23], i.e. $z_{pc}(\boldsymbol{n}) = z(\boldsymbol{n}'), \boldsymbol{n} \in \mathrm{V}_i$. T_Δ fits a surface mesh over the sensed depth point cloud (the world coordinates of the sensed depth), then projects the mesh vertices over the RGB sensor plane and fills all regular points under the projected triangles with the corresponding surface values. This is an approach built-in in GPU hardware. Fig. 4 illustrates the sensing topology with relative sizes of ToF and color pixels and their relative positions and tiling after reprojection. The first experiment considers either an upsampling only or a recursive reconstruction (two iterations tested) for different input spatial resolutions of the sensed depth.

The results are shown in Fig. 6a. As seen from the figure, the combination of super-pixel segmentation and bilateral reconstruction demonstrates a superior performance over the whole range of tested resolutions. It is also faster than the Voronoi cells fitting.

In the second experiment, we test the performance of the upsampling methods in the presence of noise. We model the noise according to the ToF principle of operation (Section 3.1) for varying noise levels γ_d = [0.25, 0.5, 0.75, 1, 2, 2.5, 3] m:

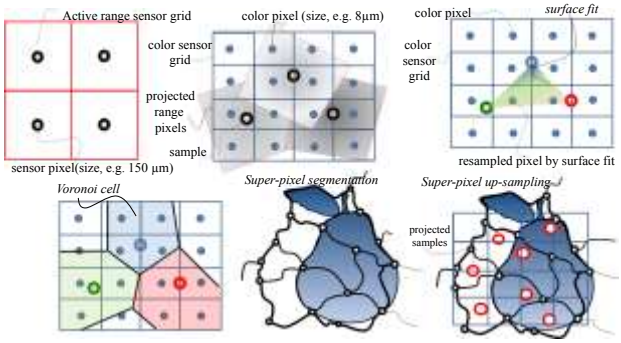

Fig 4. Non-uniform upsampling for projected data in asymmetric non-confocal RGB+ToF setup. Top row: relative size of ToF sensors, same projected on the RGB sensor plane, surface fitting through triangulation of depth pixels; Bottom row: Voronoi cell fitting around depth pixels, super-pixel segmentation and relative positions of depth samples (in red).

$$d_{noisy}(\boldsymbol{m}) = d(\boldsymbol{m}) + \sqrt{\gamma_d (d(\boldsymbol{m})/d_{max})^{1.9}}\,\varepsilon(\boldsymbol{n}), \quad (6)$$

where $\varepsilon \in N(0,1)$. The noisy range map is denoised by applying the method proposed in [21]. Both the noisy and denoised range maps are projected on the color image plane to get noisy and denoised projected depth maps on non-uniform grid. The sole upsampling and the combination of upsampling and reconstruction are applied to both the noisy and denoised range maps. The results for the three upsampling methods are shown in Figure 6 b, c, d. As seen in the figures, the surface fit upsampling method is inferior. It is more prone to distortions as the noise affects the position of the mesh vertices and the surface triangles might vary substantially. This is more difficult to correct with the bilateral filter. The depth edges appear smeared as a result of the wide and non-spatially adaptive surface triangles. Low to moderate noise levels do not require preliminary denoising. The regularized reconstruction is good enough to cope with such amount of noise. While the super-pixel and Voronoi cell based upsampling show comparable results, the former is faster than the latter. While the Voronoi cells are determined by the topological position of the depth pixels only, the spatially-adaptive nature of the super-pixels allows reducing the number of depth pixels in the sparse depth representation. In our experiments, the number of super-pixel segments was set effectively twice lower than the number of sensed (and projected) depth pixels. This offers even more compact depth representation. For all methods, the regularized reconstruction of denoised depth (green curves) shows better consistency, i.e. the curves are like a sheaf. Visual appearance of the three depth reconstruction methods is exemplified in Figure 5.

5. CONCLUSIONS

In this work, we modified a recently proposed asymmetric view + depth 3D scene representation to work for RGB-Z data sensed by a combination of color and ToF sensors. In

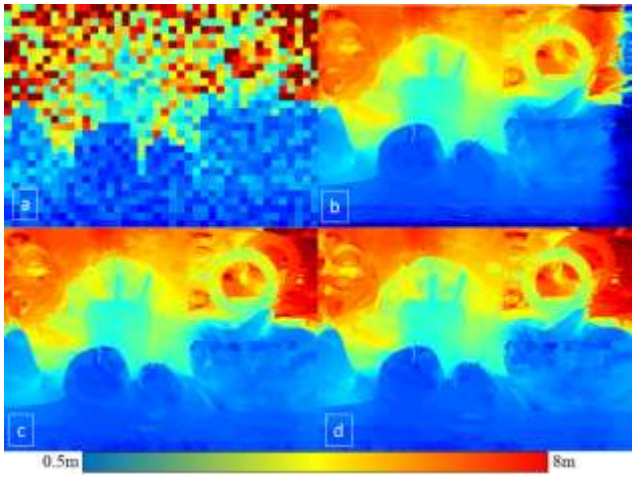

Fig 5. Input depth map and regularized outputs a) Depth sensed by sensor of 40x30 pixels, pixel size of 800 μm and σ_D^2 = 3 m; Regularized reconstruction by b) Surface-fit; c) Voronoi cells; d) Super-pixels.

essence, we propose keeping the depth in low-resolution on a plane aligned with the color image. In our concept, the correlation between the two modalities is made explicit through a combination of super-pixel color image segmentation and regularized depth reconstruction by a spatially adaptive filter. This leads to spatially sparsified and regular depth image, which can be further compressed by a suitable image compression technique [10]. The proposed scheme seems quite suitable for asymmetric scene sensing since the pixel size of the range sensor is bigger than the size of the color sensor and correspondingly, the resolution of the range image is already smaller than the resolution of the color one. However, several important implications should be taken into account: 1) The color and range cameras should be well calibrated in order to find the correct projections of the depth pixels' coordinates on the color image plane; 2) Since the depth is a mapping function, it should come to the scheme reasonably-well denoised – a high amount of noise invalidates the scheme while low to moderate noise levels are tackled well by the regularized reconstruction; 3) Super-pixel based upsampling works well and is faster than the nearest neighbor upsampling based on Voronoi cells, which reflects only the spatial positions of the measured depth samples. 4) Surface fit, while considered attractive bacause it is available on GPU hardware, is more prone to noise. A preliminary denoising in the point cloud might be considered as a remedy [21].

10. REFERENCES

[1] K. Mueller, P. Merkle, T. Wiegand, "3-D Video Representation Using Depth Maps," *Proceedings of the IEEE*, vol. 99 (4), pp. 643–656, April, 2011.

[2] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Computer Vision*, vol.47(1/2/3), pp.7–42, June 2002.

[3] D. Scharstein, R. Szeliski, "High-Accuracy Stereo Depth Maps Using Structured Light," *Proc. of CVPR*, vol.1, pp. 195-202, 2003
[4] R. Lange, P. Seitz, "Solid State ToF Range Camera," *Quantum Electronics*, vol. 37 (3), pp. 390-297, March, 2001.
[5] Y. Morvan, D. Farin, "Platelet-based coding of depth maps for the transmission of multiview images," *Proc. of SPIE*, vol. 6055, pp. 93–100, January, 2006.
[6] N. Ponomarenko, V. Lukin, A. Gotchev, K. Egiazarian, "Intra-frame Depth Image Compression Based on Anisotropic Partition Scheme and Plane Approximation," *Proc. of 2nd Int. Conference Immerscom,* 6 pages, May, 2009.
[7] I. Schiopu, I. Tabus, "Lossy depth image compression using greedy rate-distortion slope optimization," *IEEE Signal Processing Letters*, vol. 20 (11), pp. 1066–1069, 2013.
[8] P. Ndjiki-Nya, M. Koppel. D. Doshkov, H. Lakshman, K. Muller, P. Merkle, T. Wiegand, "Depth Image-based Rendering with Advanced Texture Synthesis for 3D Video," *IEEE Trans. of Multimedia*, vol.13 (3), pp. 453–465, 2011.
[9] I. Tosic, S. Drewes, "Learning Joint Intensity-depth Sparse Representations," *IEEE Trans. on Image Processing*, vol. 23 (5), pp. 2122 – 2132, April, 2014.
[10] M. Georgiev, E. Belyaev, A. Gotchev, "Depth Map Compression Using Color-driven Isotropic Segmentation and regularised reconstruction," *Proc. of Data Compression Conference 2015*, accepted.
[11] X. Ren and J. Malik, "Learning a Classification Model for Segmentation," *Proc. of 9th IEEE International Conference on Computer Vision*, vol. 1, pp. 10–17, October, 2003.
[12] M. Bergh, X. Boix, G. Roig, B. Gool, "SEEDS: Superpixels Extracted via Energy-driven Sampling," *Proc. of ECCV*, vol. 7578, pp. 13–26, October, 2012.
[13] R. Achanta, A. Shaji, K. Smith, S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-art Superpixels Methods," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34 (11), pp. 2274–2282, June, 2010.
[14] A. Buades, J. Morel, "A Non-local Algorithm for Image Denoising," *Proc. of IEEE Conference of Computer Vision and Pattern Recognition*, vol. 2, pp. 60-65, June, 2005.
[15] K. Dabov, A. Foi, V. Katkovnik, K, Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Trans. on Image Processing*, vol. 16(8), pp. 2080-2095, July, 2007.
[16] C. Tomasi, R. Manduchi, "Bilateral Filtering for Gray and Color Images," *Proc. of the IEEE International Conference on Computer Vision*, vol.1, pp. 839-847, 1998.
[17] J. Kopf, M. Cohen, D. Lischinski, M. Uyttendaele, "Joint Bilateral Upsampling," *Proc. of ACM SIGGRAPH*, vol. 26 (3), pp. 1-5, July, 2007.
[18] S. Smirnov, A. Gotchev, K. Egiazarian, "Methods for Depth-map Filtering in View+Depth 3D Video Representation," *EURASIP Journal Adv. in Signal Processing*, vol. 25 (2), 2012.
[19] M. Frank, M. Plaue, H. Rapp, U. Koethe, B. Jaehne, F. Hamprecht, "Theoretical and Experimental Error Analysis of Continuous-Wave Time-Of-Flight Range Cameras," *Optical Engineering*, vol. 48 (1), pp. 1-24, 2009.
[20] M. Georgiev, A. Gotchev, M. Hannuksela, "De-noising of Distance Maps Sensed by Time-of-flight devices in Poor Sensing Environment," *Proc. of IEEE ICASSP*, vol. 1, pp. 1533-1537, May, 2013.

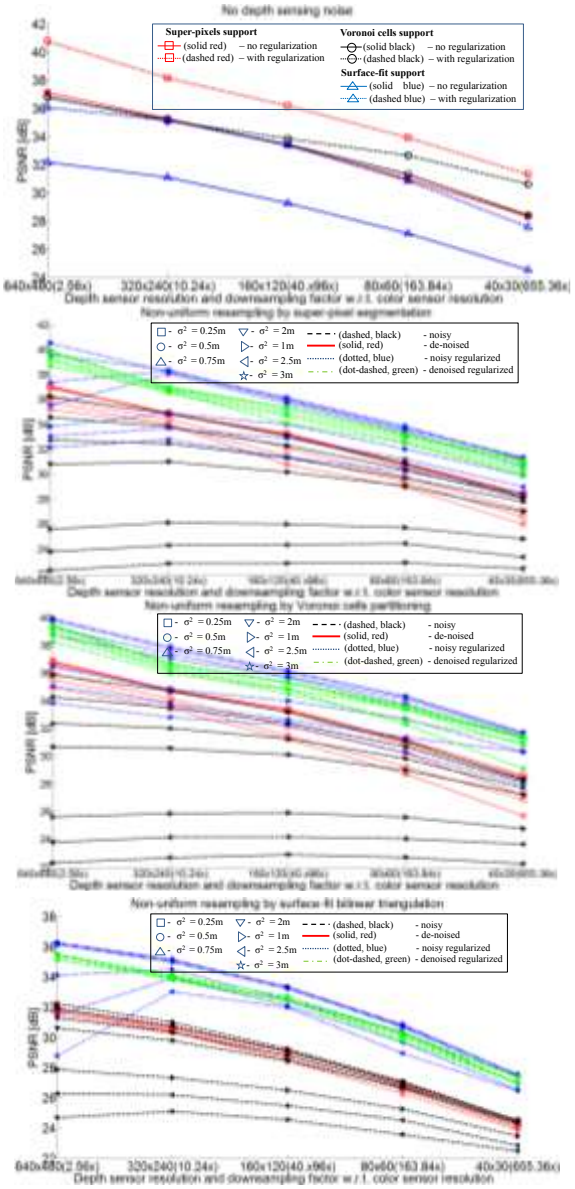

Fig 6. Performance of proposed scheme in terms of PSNR between ground-true and reconstructed depth a) noise-free reconstruction for three upsampling schemes; Reconstruction from noisy data with three upsampling schemes: b) Super-pixels, c) Voronoi cells, d) Surface-fit

[21] M. Georgiev, A. Gotchev, M. Hannuksela, "Joint De-Noising and Fusion of 2D Video and Depth Map Sequences Sensed by Low-powered ToF Range Sensor," *Proc. of IEEE Multimedia and Expo Workshops*, vol. 1, 4 pages, July, 2013.
[22] M. Georgiev, A. Gotchev, M. Hannuksela, "Real-time Denoising of Tof Measurements by Spatio-Temporal Non-Local Mean Filtering," *Proc. of IEEE Multimedia and Expo Workshops*, vol. 1, 6 pages, July 2013.
[23] T. Strohmer, "Efficient methods for digital signal and image reconstruction from Non-uniform Samples." *PhD thesis*, University of Vienna, 1993.