Precision Enhancement of 3-D Surfaces from Compressed Multiview Depth Maps

Pengfei Wan, *Student Member, IEEE*, Gene Cheung, *Senior Member, IEEE*, Philip A. Chou, *Fellow, IEEE*, Dinei Florencio, *Senior Member, IEEE*, Cha Zhang, *Senior Member, IEEE*, and Oscar C. Au, *Fellow, IEEE*

***Abstract*—Transmitting depth maps captured from multiple viewpoints of a 3-D scene enables a wide range of receiver-side 3-D applications, including virtual view synthesis via depth-image-based rendering (DIBR). Observing that compressed depth maps from different viewpoints constitute multiple descriptions (MD) of the same signal, we propose to reconstruct 3-D surfaces of the scene by considering multiple compressed depth maps jointly. Specifically, we propose an alternating projection algorithm, inspired by the theory of projection onto convex sets (POCS), which at convergence returns a 3-D surface that satisfies three sets of conditions: spatial smoothness prior, quantization bin constraints in the block transform domain, and inter-view consistency. We present a theoretical proof that shows convergence of our algorithm under benign conditions. Compared to existing multiview depth map denoising schemes and single image de-quantization schemes, our proposed solution achieves higher objective quality for both reconstructed depth maps and synthesized virtual views.**

***Index Terms*—Multiview video plus depth, precision enhancement, projection onto convex sets.**

I. Introduction

BY TRANSMITTING depth maps captured from multiple viewpoints of the same 3-D scene, a receiver can recover partial geometry of the scene, enabling a wide range of 3-D applications, including geometry modeling [1]–[3], depth-aware image processing such as matting [4] and refocusing [5], gesture recognition [6], and virtual view synthesis via *depth-image-based rendering* (DIBR) [7]. Transmitting multiple depth maps constitutes a large transmission cost, however, and thus there are recent proposals on efficient multiview video coding algorithms [8]–[13]. In this paper, we focus instead on an orthogonal problem at the decoder: how to best reconstruct 3-D surfaces in the scene given the compressed multiview depth maps?

Manuscript received January 23, 2015; accepted March 31, 2015. Date of publication April 15, 2015; date of current version April 21, 2015. This work was supported in part by the Microsoft Research CORE program. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Glenn Easley.

P. Wan and O. C. Au are with the ECE Department, Hong Kong University of Science and Technology, Hong Kong, China. (e-mail: leoman@ust.hk; eeau@ust.hk).

G. Cheung is with National Institute of Informatics, Tokyo 101-8430, Japan (e-mail: cheung@nii.ac.jp).

P. A. Chou, D. Florencio, and C. Zhang are with Microsoft Research, Redmond, WA 98052 USA (e-mail: pachou@microsoft.com; dinei@microsoft.com; chazhang@microsoft.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2423372

We propose to reconstruct 3-D surfaces by jointly considering multiple compressed depth maps of different viewpoints. The key observation is that each depth map is a unique description of the 3-D scene, which implies that multiview depth maps constitute *multiple descriptions* (MD) of the same 3-D scene. Recall that in multiple description scalar quantization (MDSQ), a scalar x is quantized via two different quantizers $\mathcal{Q}_1$ and $\mathcal{Q}_2$ into quantization bins (q-bin) $\mathcal{Q}_1(x)$ and $\mathcal{Q}_2(x)$ with corresponding quantization indices (q-index) transmitted to the receiver. At the receiver, if both q-indices are received, the receiver can conclude that the target signal resides in the intersection of the two q-bins, *i.e.*, $x \in \mathcal{Q}_1(x) \cap \mathcal{Q}_2(x)$. This results in a higher-precision reconstructed signal than the case where only one q-index is received. Analogously, in our case where compressed multiview depth maps of the same 3-D scene are received at the decoder, we seek to reconstruct a high-precision 3-D surface which is inside the intersection of the quantized descriptions of the multiple views.

By 3-D surface we mean its chosen representation, which is the set of multiview depth maps in this paper. So reconstructing a high-precision 3-D surface is formulated as reconstructing high-precision multiview depth maps. Inspired by the theory of *projection onto convex sets* (POCS) [14], [15] and its applications to signal recovery [16]–[19], we propose an alternating projection algorithm to reconstruct multiview depth maps. In the case of two views, an estimated left depth map is first warped to a depth map as observed from the right view via DIBR. A q-bin projection (to ensure the reconstructed depth signal lies inside q-bins) and spatial filter (to ensure the reconstructed depth map agrees with spatial smoothness prior) are then applied in order. The updated right view depth map is then warped back to a depth map as observed from the left view again for q-bin projection and spatial filtering. The alternating steps terminate when the computed depth maps converge in both views. Our proposed algorithm can be applied more generally to any number of views, as explained in Section IV-B. Experiments show that our algorithm outperforms existing multiview depth map denoising algorithms and single image de-quantization schemes in objective quality of the reconstructed multiview depth maps as well as DIBR-synthesized virtual views.

This letter is organized as follows. We first overview the related works and the system of our proposal in Section II and III respectively. Then we present our proposed algorithm and justification in Section IV and V. Experiment results and conclusions are shown in Section VI and VII, respectively.

II. Related Work

There are recent proposals that jointly denoise multiview depth maps corrupted by acquisition or estimation noise

[20]–[26], where inter-view consistency among multiview depth maps is considered as an additional constraint for depth map denoising. *The problem we are addressing, however, is not denoising of noise-corrupted depth maps, but de-quantization of compressed depth maps*. In our problem, transform coefficients of a depth block are mapped to q-bin indices during quantization in lossy compression, resulting in deterministic q-bin constraints: a reconstructed coefficient must live inside the q-bin designated by the transmitted q-index. We will show in our experiments that multiview depth map denoising algorithms, when applied naïvely to our problem setting, are inferior to our proposal that is specialized for de-quantization of compressed multiview depth maps.

There are also works that de-quantize a single block-transform-coded image to alleviate blocking artifacts [17], [18]. In our problem, since multiview depth signals describe the same 3-D scene, our proposed method considers both inter-view consistency and q-bin constraints, leading naturally to a POCS-inspired algorithm for joint reconstruction of multiview depth maps. We will show in our experiments that our proposal outperforms single image de-quantization schemes.

III. System Overview

For simplicity, we consider a scenario where the 3-D surface is represented by two depth maps (left and right), which are compressed by a conventional block-based codec (*e.g.*, H.264 [27], JPEG [28]). Each (residual) block is transformed to frequency domain via a fixed transform such as *Discrete Cosine Transform* (DCT), and the resulting coefficients are quantized and entropy-coded. At the decoder, we seek to reconstruct two depth maps with higher precision by jointly considering quantized transform coefficients of both maps. We assume that the acquisition noise in pre-compressed depth maps is sufficiently small that the captured 3-D surface live in the non-empty intersections of q-bins in the two descriptions.

Let $\mathbf{d}_l^o$ denote the original pre-compressed left depth map of resolution $H \times W$. The transform coefficients are $\mathbf{c}_l^o = T(\mathbf{d}_l^o - \mathbf{d}_l^{\text{pre}})$, where $T(\cdot)$ is the block-based transform operator, and $\mathbf{d}_l^{\text{pre}}$ is the predictor. In a conventional decoder that decodes each view separately, the reconstructed left depth map is $\mathbf{d}_l^{\text{anc}} = T^{-1}(\mathbf{c}_l^q) + \mathbf{d}_l^{\text{pre}}$, where $\mathbf{c}_l^q$ denotes the q-bin centers of quantized coefficients. Symbols for the right view can be similarly defined. In this work, reconstructed left and right depth maps are denoted by $\mathbf{d}_l^{\text{opt}}$ and $\mathbf{d}_r^{\text{opt}}$. The inputs to our problem are: i) $H \times W$ matrices of quantized transform coefficients $\mathbf{c}_l^q$ and $\mathbf{c}_r^q$ for the left and right views, and ii) $H \times W$ matrix of q-step size $\boldsymbol{\kappa}$ (same for both views).

IV. Proposed Algorithm

A. Problem Formulation

We formulate our multiview depth map de-quantization problem with the following signal prior and constraints.

1) Spatial Smoothness Prior: Unlike color images, depth images do not capture textural contents of objects, and thus are known to be *piecewise smooth* (PWS) [29]–[31]. PWS here means that while depth images contain sharp edges (*e.g.*, contours that outline shapes of foreground objects), surfaces away from edges are slow-varying in space.

To enforce a PWS prior in images, one can apply an edge-adaptive low-pass filter to eliminate high-frequency components that are not sharp edges. As in [2], in this work we use the well-known bilateral filter [32]–[34] which preserves edges and suppresses strong quantization noises well for depth maps.

2) Quantization Bin Constraints: We require the transform coefficients of the reconstructed depth maps to fall within the q-bins designated by the transmitted q-indices. Mathematically, we write for each coefficient index i:

$$\mathbf{c}(i) \in \left[\mathbf{c}^q(i) - \frac{\boldsymbol{\kappa}(i)}{2}, \mathbf{c}^q(i) + \frac{\boldsymbol{\kappa}(i)}{2}\right), \forall i \tag{1}$$

To enforce q-bin constraints (1), one can simply clip each coefficient by the corresponding q-bin boundaries:

$$\min\left(\mathbf{c}^q(i) + \frac{\boldsymbol{\kappa}(i)}{2}, \max\left(\mathbf{c}^q(i) - \frac{\boldsymbol{\kappa}(i)}{2}, \mathbf{c}(i)\right)\right), \forall i \tag{2}$$

3) Inter-view Consistency: This constraint means that the warping of the left reconstructed depth map to the right view must be consistent with the right reconstructed depth map, and vice versa. Let $w_{l\to r}(\cdot)$ be the warping operator[1] that performs DIBR to transpose a left depth map to the right view. Because of disocclusion and out-of-view problems, not every 3-D voxel observable in the right view is visible from the left view, and the warped image $w_{l\to r}(\mathbf{d}_l)$ will contain holes (*i.e.*, $w_{l\to r}(\mathbf{d}_l)(i) < 0$ for some pixels i). Thus for inter-view consistency, we only require the available valid pixels in the warped view $w_{l\to r}(\mathbf{d}_l)$ to match the right depth map $\mathbf{d}_r$:

$$\begin{aligned} &|w_{l\to r}(\mathbf{d}_l)(i) - \mathbf{d}_r(i)| \le \epsilon, \quad \forall\{i | w_{l\to r}(\mathbf{d}_l)(i) \ge 0\} \\ &|\mathbf{d}_l(i) - w_{r\to l}(\mathbf{d}_r)(i)| \le \epsilon, \quad \forall\{i | w_{r\to l}(\mathbf{d}_r)(i) \ge 0\} \end{aligned} \tag{3}$$

Note that we require a pixel match in (3) to be within a threshold, because rounding and interpolation operations performed after warping to ensure each warped pixel lands on the 2D image grid will inherently introduce errors [21].

Our goal is to construct left and right depth maps $\mathbf{d}_l^{\text{opt}}$ and $\mathbf{d}_r^{\text{opt}}$ that satisfy the above signal prior and constraints simultaneously. We describe our reconstruction algorithm next.

B. Proposed Alternating Projection Algorithm

To de-quantize multiview depth maps that represent a 3-D surface with enhanced precision, we propose a POCS-inspired alternating projection algorithm shown in Algorithm 1.

At iteration k, step 2 of Algorithm 1 is to obtain $\mathbf{d}_r^{(k)}$ from $\mathbf{d}_l^{(k-1)}$. Specifically, we first warp $\mathbf{d}_l^{(k-1)}$ to the right view by operator $w_{l\to r}(\cdot)$. Holes in $\mathbf{d}_r^{(k)}$ are then filled using available $\mathbf{d}_r^{(k-1)}$. For a pixel i (at column $\text{col}(i)$) in the non-hole regions, $\mathbf{d}_r^{(k)}(i)$ is interpolated by taking the average of depth values of the pixels in $\mathbf{d}_l^{(k-1)}$ whose warped pixels locate at columns (off 2D pixel grid) in range $[\text{col}(i) - 0.5, \text{col}(i) + 0.5)$, and whose pixel values are in range $[\mathbf{d}_r^{(k-1)}(i) - \tau, \mathbf{d}_r^{(k-1)}(i) + \tau]$. Constant τ is a threshold to reject outliers (*e.g.*, foreground pixels that wrongly warped to the background, and vice versa).

Algorithm 1 is applicable to scenarios when there are $N > 2$ multiview depth maps. To assure small distortions, our view warping direction is as follows: one iteration of our algorithm

[1]For rectified views, warping operator simply translates a pixel in the left view to a horizontally shifted location in the right view and vice versa [35].

contains the view warping from view 1 to 2, then 2 to 3 etc. until N, then backwards from view N to $N-1$, etc.

Algorithm 1

1: Initialize depth maps $\mathbf{d}_l^{(0)} = T^{-1}(\mathbf{c}_l^q) + \mathbf{d}_l^{\text{pre}}$, $\mathbf{d}_r^{(0)} = T^{-1}(\mathbf{c}_r^q) + \mathbf{d}_r^{\text{pre}}$. Iteration index $k = 1$.
2: Warp left depth map $\mathbf{d}_l^{(k-1)}$ to right view, obtaining $\mathbf{d}_r^{(k)}$ by average interpolation and hole-filling.
3: Perform block-DCT, and clip coefficients $\mathbf{c}_r = T(\mathbf{d}_r^{(k)} - \mathbf{d}_r^{\text{pre}})$ to inside designated q-bins via (2).
4: Inverse-transform back to pixel domain $\mathbf{d}_r^{(k)} = T^{-1}(\mathbf{c}_r) + \mathbf{d}_r^{\text{pre}}$, and apply bilateral filtering.
5: Warp right depth map $\mathbf{d}_r^{(k)}$ to left view, obtaining $\mathbf{d}_l^{(k)}$ by average interpolation and hole-filling.
6: Perform block-DCT, and clip coefficients $\mathbf{c}_l = T(\mathbf{d}_l^{(k)} - \mathbf{d}_l^{\text{pre}})$ to inside designated q-bins via (2).
7: Inverse-transform back to pixel domain $\mathbf{d}_l^{(k)} = T^{-1}(\mathbf{c}_l) + \mathbf{d}_l^{\text{pre}}$, and apply bilateral filtering.
8: $k = k + 1$, repeat 2-7 until $\mathbf{c}_l$ and $\mathbf{c}_r$ converge.

One iteration of Algorithm 1 is composed of view warping, block-based DCT, coefficient clipping, block-based inverse DCT, and bilateral filtering for N views. Complexity of view warping and coefficient clipping is $O(HW)$. Block-based DCT and inverse DCT are essentially matrix multiplication, so the complexity is $\frac{H}{B}\frac{W}{B}O(B^3) = O(HW)$ where B is the block size. We adopt the constant-time bilateral filter [33], [34] whose complexity is also $O(HW)$. Thus the complexity for one iteration is $O(NHW)$ which is linear to the number of pixels. Algorithm convergence is proven in the next section.

V. Convergence Proof

We now prove the convergence of Algorithm 1 for a simpler case where we consider one row of pixels on the epipolar plane of two rectified views. Now $\mathbf{d}_l, \mathbf{d}_r \in \mathbb{R}^W$ are 1D signals in Hilbert space $\mathcal{H}$. To begin with, we make two assumptions.

Assumption 1: 1D signals $\mathbf{d}_l$ and $\mathbf{d}_r$ are bandlimited, *i.e.*, their non-zero frequency components are no larger than $\omega = 2\pi \cdot \frac{1}{2T} = \frac{\pi}{T}$, where T is the sampling interval.

This means that the corresponding continuous signals $\mathbf{d}_l^c$ and $\mathbf{d}_r^c$ can be perfectly reconstructed via Whittaker-Shannon interpolation formula: $\mathbf{d}_l^c(x) = \sum_{n=1}^{N} \mathbf{d}_l[n]\ \text{sinc}(\frac{x-nT}{T})$. When ω is small enough, this also means that no points on left signal $\mathbf{d}_l^c(x)$ is occluded in the right view since self-occlusion occurs when the signal gradient is unbounded.

Assumption 2: the view warping processes can be approximated as fixed linear transforms.

This means that left and right depth maps are related by a $W \times W$ matrix $\mathbf{M}$; *i.e.*, $\mathbf{d}_r = \mathbf{M}\mathbf{d}_l$ (here we only consider the 3-D voxels which are visible in both views).

Based on the assumptions, we next prove that Algorithm 1 essentially cyclically projects variable $\mathbf{d}_l$ onto the following 4 closed convex sets in $\mathcal{H}$: 1) $\mathcal{S}_l^q$: the set of $\mathbf{d}_l$ whose DCT coefficients $\mathbf{c}_l = T(\mathbf{d}_l - \mathbf{d}_l^{\text{pre}})$ are within the q-bin constraints in the left view; 2) $\mathcal{S}_r^q$: the set of $\mathbf{d}_l$ whose DCT coefficients $\mathbf{c}_r = T(\mathbf{M}\mathbf{d}_l - \mathbf{d}_r^{\text{pre}})$ are within the q-bin constraints in the right view; 3) $\mathcal{S}_l^\omega$: the set of $\mathbf{d}_l$ that are bandlimited by frequency ω; and 4) $\mathcal{S}_r^\omega$: the set of $\mathbf{d}_l$ satisfying $\mathbf{M}\mathbf{d}_l$ are bandlimited by ω. According to [15], cyclic projections in Hilbert space are guaranteed to converge to the intersection of convex sets. Therefore to prove the convergence of Algorithm 1, we only need to prove the following propositions:

1) $\mathcal{S}_l^q$, $\mathcal{S}_r^q$, $\mathcal{S}_l^\omega$ and $\mathcal{S}_r^\omega$ are convex sets for $\mathbf{d}_l$.
2) Mapping onto $\mathcal{S}_l^q$ ($\mathcal{S}_r^q$) is a projection in $\mathcal{H}$.
3) Mapping onto $\mathcal{S}_l^\omega$ ($\mathcal{S}_r^\omega$) is a projection in $\mathcal{H}$.

Proof: Each q-bin constraint for a DCT coefficient is convex, and jointly considering a set of convex constraints also leads to a convex set $\mathcal{S}_l^q$ for $\mathbf{d}_l$. For $\mathcal{S}_r^q$, due to the linear transform, a convex set for $\mathbf{d}_r$ is also a convex set for $\mathbf{d}_l$.

Consider two signals $\mathbf{d}_l^1$ and $\mathbf{d}_l^2$ that are bandlimited by frequency ω. Clearly a convex combination $\mathbf{d}_l = \lambda\mathbf{d}_l^1 + (1-\lambda)\mathbf{d}_l^2$ is also bandlimited by frequency ω. Hence $\mathcal{S}_l^\omega$ is a convex set for $\mathbf{d}_l$. Likewise $\mathcal{S}_r^\omega$ is a convex set for $\mathbf{d}_r$. Because of linear transform, $\mathcal{S}_r^\omega$ is also a convex set for variable $\mathbf{d}_l$.

Block-DCT coefficient clipping in the left view is a projection in space $\mathcal{H}$, as the minimum change in energy (distance) is induced to bring $\mathbf{d}_l$ to inside convex set $\mathcal{S}_l^q$. So step 6 is a projection. Due to uniform rotation of the space, right view coefficient clipping (step 3) is also a projection in $\mathcal{H}$.

A low-pass filter that removes only high frequency energy (those over ω) achieves minimum distance to $\mathcal{S}_l^\omega$, and hence is a projection in $\mathcal{H}$, so step 7 is a projection. Due to uniform rotation of the space, step 4 is also a projection in $\mathcal{H}$. ■

VI. Experiments

To reconstruct the compressed multiview depth maps, we employ the following schemes for comparison: 1) ANC: anchor method that decodes depth maps separately, with output $\mathbf{d}^{\text{anc}}$; 2) BLF: direct bilateral filtering on the separately decoded depth maps $\mathbf{d}^{\text{anc}}$; 3) DBLK: a single image de-blocking method [17], applied on each compressed depth map separately; 4) JVDF: joint–view depth filtering proposed in [20] that denoises multiview depth maps by improving inter-view consistency; 5) IVDC: inter-view depth consistency test and enhancement [23], [24], which has a similar idea with JVDF except that it iteratively denoises multiview depth maps from a statistical perspective; 6) PROP: our proposed method, with output $\mathbf{d}^{\text{opt}} = T^{-1}(\mathbf{c}^{(k^*)}) + \mathbf{d}^{\text{pre}}$ (Algorithm 1 converged at iteration k^*); and 7) PROP-1V: our proposed method without view warping, which can be viewed as the modified version of DBLK using bilateral filter instead of Gaussian filter.

Test sequences include `dude` (480×800) consisting of multiview depth maps of a synthesized human model [36]; `new tsukuba` (480×640) synthesized stereo depth maps [37]; as well as `bowling` and `aloe` (368×416) which are stereo depth maps of natural scenes [38]. For conformity, all test images were 8-bit disparity maps (depth values can be calculated from disparity values). Test images were compressed using JPEG (Quality Factor $= 25, 50, 75$) and H.264 (Quantization Parameter $= 35$). For H.264 compression of multiview depth maps, we alternated across views to either intra-code the original disparity image as I-frame, or intra-code the difference image—a pre-computed difference image between target image and the predictor image constructed via warping from adjacent

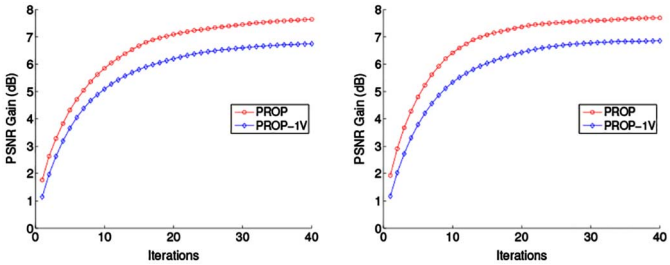

Fig. 1. Iterations vs. depth PSNR gain over ANC in experiment "JPEG QF $=$ 50" of Table 1(a). Two sub-figures correspond to the left and right views.

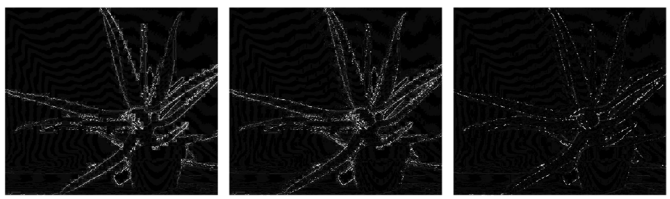

Fig. 2. Left-view depth error maps in experiment "JPEG QF $=$ 50" of Table 1(d). From left to right: method ANC, IVDC, PROP.

independently coded images. 3×3 bilateral filter was used in the experiments. We set $\tau = 10$ for view interpolation.

Algorithm 1 terminates when the mean-absolute-difference between iterations is smaller than 10^{-8} for both $\mathbf{c}_l$ and $\mathbf{c}_r$, or the iteration number reaches the maximal allowed value (40 in our experiments). As shown in Fig. 1, we observed convergence of proposed Algorithm 1 for each view in around 30 iterations. Similar convergence behaviors were observed in both the JPEG and H.264 experiments.

We measure the performance of different methods using the *Peak Signal-to-Noise Ratio* (PSNR) of both the reconstructed multiview depth maps and their synthesized color views[2]. Uncompressed depth maps and the corresponding synthesized views served as the ground-truth. Table I summarizes the numerical results. In terms of quality of reconstructed multiview depth maps, we see our PROP, PROP-1V achieved the best performance. The reason BLF performed poorly is because it does not consider information from multiple views. JVDF and IVDC performed poorly because they are designed for denoising, not de-quantization. DBLK performed the worst because it uses low-pass filter rather than edge-preserving filter so important depth edges are blurred. Comparing PROP-1V and PROP, we verified that combining information from multiple viewpoints reduced the distortion of reconstructed depth maps. The PSNR results of synthesized virtual views using the reconstructed depth maps in Table I also verified the effectiveness of PROP and PROP-1V. The reason is that our proposed method strongly suppressed quantization distortions along object edges (demonstrated in Fig. 2), so the synthesized view naturally exhibited smaller color distortion.

Table II shows the PSNR of a depth map reconstructed by PROP under varying baselines and number of views for `dude`. θ is the view angle between adjacent cameras (baseline $\propto \tan(\theta/2)$). Comparing results for different baselines, we see that the PSNR results increased then decreased, as baseline increased from zero. This was because if the camera distance was too small, then multiview depth maps provided almost identical information of the 3-D scene; on the other hand, large baseline led to large positional errors in view warping procedure which degraded the performance. Comparing results under different number of viewpoints, we see that in general PSNR results increased as more views were involved. This was because more views were likely to introduce more new information of the 3-D surface, thus multiview depth maps were reconstructed with higher precision.

[2]We exclude hole pixels from synthesized view PSNR calculation as they account for a very small portion of pixels.

TABLE I
PSNR RESULTS IN DB. FOR EACH TABLE CELL, THE UPPER VALUE IS THE AVERAGE PSNR OF THE RECONSTRUCTED MULTIVIEW DEPTH MAPS, THE LOWER VALUE IS THE PSNR OF THE CORRESPONDING SYNTHESIZED COLOR IMAGE AT THE CENTER VIEW. (A) `dude` (2 VIEWS, ANGLE $= 5°$) (B) `new tsukuba` (2 VIEWS) (C) `bowling` (2 VIEWS) (D) `aloe` (2 VIEWS)

	ANC	BLF	DBLK	JVDF	IVDC	PROP-1V	PROP
JPEG	33.06	33.63	32.64	33.36	33.75	36.15	**36.66**
QF=25	29.39	29.50	28.69	29.12	29.89	32.99	**33.33**
JPEG	35.36	36.39	34.88	35.71	36.19	42.16	**43.03**
QF=50	29.99	30.82	29.66	30.11	30.56	34.77	**34.77**
JPEG	38.71	40.97	37.64	39.38	39.82	46.81	**47.12**
QF=75	31.16	32.16	30.71	31.35	31.72	**36.02**	35.60
H.264	36.94	37.91	35.56	37.53	38.20	42.05	**42.32**
QP=35	30.78	31.21	29.95	30.96	31.29	34.02	**34.92**

(a)

	ANC	BLF	DBLK	JVDF	IVDC	PROP-1V	PROP
JPEG	34.50	34.72	33.91	34.59	34.44	36.02	**36.68**
QF=25	29.04	29.17	28.37	28.82	29.55	30.20	**30.49**
JPEG	37.22	37.62	36.38	37.37	37.40	40.03	**41.22**
QF=50	29.84	30.02	28.93	29.69	30.53	31.33	**31.89**
JPEG	40.67	41.54	39.51	40.98	40.54	47.25	**47.61**
QF=75	30.54	30.93	29.59	30.70	31.24	32.83	**33.01**
H.264	38.26	38.60	36.79	38.57	38.59	41.01	**41.33**
QP=35	30.15	30.37	29.22	29.82	30.88	31.91	**32.76**

(b)

	ANC	BLF	DBLK	JVDF	IVDC	PROP-1V	PROP
JPEG	36.42	36.72	35.44	36.77	36.84	37.95	**38.33**
QF=25	31.54	31.54	31.25	31.45	31.81	32.62	**33.13**
JPEG	38.83	39.29	37.89	39.23	39.35	41.59	**42.33**
QF=50	31.95	32.21	31.88	32.11	32.30	33.45	**34.00**
JPEG	41.58	42.51	40.57	42.16	42.27	47.51	**48.26**
QF=75	32.54	32.81	32.66	32.38	32.67	34.40	**34.68**
H.264	38.71	39.02	37.64	39.51	39.86	41.30	**41.57**
QP=35	31.85	32.04	31.89	32.09	32.26	33.26	**33.74**

(c)

	ANC	BLF	DBLK	JVDF	IVDC	PROP-1V	PROP
JPEG	33.72	34.01	33.05	34.03	33.74	35.09	**35.40**
QF=25	27.07	27.27	26.56	27.14	27.33	27.95	**28.33**
JPEG	35.64	36.12	35.11	36.08	35.67	38.49	**39.10**
QF=50	27.60	28.02	27.33	27.88	28.17	29.04	**29.37**
JPEG	38.22	39.18	37.58	38.92	38.35	44.79	**45.10**
QF=75	28.14	28.51	27.88	28.29	28.59	29.62	**29.89**
H.264	36.06	36.40	34.85	36.68	36.37	38.86	**39.15**
QP=35	27.64	27.79	27.01	27.76	28.02	28.74	**29.13**

(d)

TABLE II
DEPTH PSNR OF A FIXED VIEW OF DUDE (JPEG QF $=$ 50)

	2 views	3 views	5 views	7 views
$\theta = 0°$	42.42	42.42	42.42	42.42
$\theta = 1.5°$	43.01	43.41	43.61	43.77
$\theta = 3°$	43.03	43.77	43.70	43.95
$\theta = 5°$	43.48	43.97	44.22	44.37
$\theta = 10°$	43.03	43.63	43.72	43.68
$\theta = 20°$	43.02	43.34	43.48	43.48

VII. CONCLUSION

Observing that depth maps captured from different viewpoints are actually different descriptions of the same 3-D scene, we propose to de-quantize multiview depth maps by jointly considering their compressed versions, so that specified uncertainties in multiple views are considered simultaneously to enhance the overall precision of the represented 3-D surface. Experiments verify the effectiveness of our proposed 3-D surface precision enhancement method.

This paper previously published in IEEE Signal Processing Letters

REFERENCES

[1] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proc. 23rd Annu Conf. Computer Graphics and Interactive Techniques*, 1996, pp. 303–312, ser. SIGGRAPH '96.

[2] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proc. 2011 10th IEEE Int Symp Mixed and Augmented Reality*, 2011, pp. 127–136, ser. ISMAR '11.

[3] Y. Gao, G. Cheung, T. Maugey, P. Frossard, and J. Liang, "3d geometry representation using multiview coding of image tiles," in *2014 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 6157–6161.

[4] W. Sun, O. C. Au, L. Xu, and Z. Yu, "Adaptive depth map assisted matting in 3d video," in *Proc. 2011 IEEE Int. Conf. Multimedia and Expo*, 2011, pp. 1–6, ser. ICME '11.

[5] J. Dorsey, S. Xu, G. Smedresman, H. Rushmeier, and L. McMillan, "Towards digital refocusing from a single photograph," in *15th Pacific Conf. Computer Graphics and Applications, 2007 PG '07*, Oct. 2007, pp. 363–372.

[6] J. Wang, Z. Liu, and Y. Wu, *Human Action Recognition with Depth Cameras*. Berlin, Germany: Springer, 2014, ser. SpringerBriefs in Computer Science.

[7] D. Tian, P.-L. Lai, P. Lopez, and C. Gomila, "View synthesis techniques for 3-D video," *Applications of Digital Image Processing XXXII, Proc. SPIE*, vol. 7443, no. 2009, pp. 74 430T–74 430T–11, 2009.

[8] D. Florencio and C. Zhang, "Multiview video compression and streaming based on predicted viewer position," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing, 2009 ICASSP 2009*, 2009, pp. 657–660.

[9] G. Cheung, V. Velisavljevic, and A. Ortega, "On dependent bit allocation for multiview image coding with depth-image-based rendering," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3179–3194, Nov. 2011.

[10] M. Magnor, P. Eisert, and B. Girod, "Multiview image coding with depth maps and 3d geometry for prediction," in *Proc. SPIE*, 2000, vol. 4310, pp. 263–271.

[11] E. Ekmekcioglu, S. Worrall, and A. Kondoz, "A temporal subsampling approach for multiview depth map compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1209–1213, Aug. 2009.

[12] C. Lee and Y.-S. Ho, "A framework of 3-D video coding using view synthesis prediction," in *2010 Picture Coding Symp.*, Krakow, Poland, May 2012.

[13] I. Daribo, G. Cheung, T. Maugey, and P. Frossard, "RD optimized auxiliary information for inpainting-based view synthesis," in *3-DTV-Conf.*, Zurich, Switzerland, Oct. 2012.

[14] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, Sep. 1996.

[15] H. H. Bauschke, J. M. Borwein, and A. S. Lewis, "On the method of cyclic projections for convex sets in hilbert space," *Contemporary Mathematics*, 1994.

[16] D. C. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1theory," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 81–94, 1982.

[17] A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, no. 1, pp. 91–95, Mar. 1992.

[18] Y. Yang, N. Galatsanos, and A. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 6, pp. 421–432, Dec 1993.

[19] P. Chou, S. Mehrotra, and A. Wang, "Multiple description decoding of overcomplete expansions using projections onto convex sets," in *Proc. Data Compression Conference*, 1999, pp. 72–81.

[20] R. Li, D. Rusanovskyy, M. Hannuksela, and H. Li, "Joint view filtering for multiview depth map sequences," in *IEEE Int. Conf. Image Processing*, Orlando, FL, Oct. 2012.

[21] W. Sun, G. Cheung, P. Chou, D. Florencio, C. Zhang, and O. Au, "Rate-constrained 3d surface estimation from noise-corrupted multiview depth videos," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3138–3151, Jul. 2014.

[22] E. Ekmekcioglu, V. Velisavljevic, and S. Worrall, "Content adaptive enhancement of multi-view depth maps for free viewpoint video," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 2, pp. 352–361, Apr. 2011.

[23] P. Rana and M. Flierl, "Depth consistency testing for improved view interpolation," in *2010 IEEE Int. Workshop on Multimedia Signal Processing (MMSP)*, Oct. 2010, pp. 384–389.

[24] P. Rana, J. Taghia, and M. Flierl, "Statistical methods for inter-viewdepth enhancement," in *3-DTV-Conf.: The True Vision - Capture, Transmission and Display of 3-D Video (3-DTV-CON), 2014*, Jul. 2014, pp. 1–4.

[25] M. Kurc, O. Stankiewicz, and M. Domanski, "Depth map inter-view consistency refinement for multiview video," in *Picture Coding Symp. (PCS), 2012*, May 2012, pp. 137–140.

[26] N. Stefanoski, C. Bal, M. Lang, O. Wang, and A. Smolic, "Depth estimation and depth enhancement by diffusion of depth features," in *2013 20th IEEE Int. Conf. Image Processing (ICIP)*, Sep. 2013, pp. 1247–1251.

[27] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[28] G. K. Wallace, "The jpeg still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[29] G. Shen, W.-S. Kim, S. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," in *IEEE Picture Coding Symp.*, Nagoya, Japan, Dec. 2010.

[30] W. Hu, G. Cheung, X. Li, and O. Au, "Depth map compression using multi-resolution graph-based transform for depth-image-based rendering," in *IEEE Int. Conf. Image Processing*, Orlando, FL, USA, Sep. 2012.

[31] W. Hu, G. Cheung, A. Ortega, and O. Au, "Multi-resolution graph fourier transform for compression of piecewise smooth images," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 419–433, Jan. 2015.

[32] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Computer Vision*, Bombay, India, 1998.

[33] K. N. Chaudhury, D. Sage, and M. Unser, "Fast $o(1)$ bilateral filtering using trigonometric range kernels," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3376–3382, Dec. 2011.

[34] K. Chaudhury, "Acceleration of the shiftable mbiO(1) algorithm for bilateral filtering and nonlocal means," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1291–1300, Apr. 2013.

[35] K. Muller, P. Merkle, and T. Wiegand, "3-d video representation using depth maps," *Proc. IEEE*, vol. 99, no. 4, pp. 643–656, Apr. 2011.

[36] Microsoft, "XNA Skinned Model Sample" [Online]. Available: http://xbox.create.msdn.com/en-US/education/catalog/sample/skinned_model, Jan. 2015, accessed

[37] S. Martull, M. Peris, and K. Fukui, "Realistic cg stereo image dataset with ground truth disparity maps," in *ICPR Workshop TrakMark2012*, 2012, vol. 111, no. 430, pp. 117–118.

[38] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE Conf. Computer Vision and Pattern Recognition, 2007 CVPR'07*, 2007, pp. 1–8, IEEE.

This paper previously published in IEEE Signal Processing Letters