DEPTH MAP COMPRESSION VIA COMPRESSED SENSING

Michel Sarkis and Klaus Diepold

Institute for Data Processing, Technische Universität München
Arcisstr. 21, 80290 Munich, Germany
sarkis@mytum.de, kldi@tum.de

ABSTRACT

We propose in this paper a new scheme based on compressed sensing to compress a depth map. We first subsample the entity in the frequency domain to take advantage of its compressibility. We then derive a reconstruction scheme to recover the original map from the subsamples using a non-linear conjugate gradient minimization scheme. We preserve the discontinuities of the depth map at the edges and ensure its smoothness elsewhere by incorporating the Total Variation constraint in the minimization. The results we obtained on various test depth maps show that the proposed method leads to lower error rate at high compression ratio when compared to standard image compression techniques like JPEG and JPEG 2000.

Index Terms—Conjugate gradient methods, image coding, image representation, stereo vision

1. INTRODUCTION

Depth from stereo is one of the most developed research areas in computer vision. Its aim is to determine an accurate representation of a scene in 3D by computing a disparity (depth) map of the corresponding stereo image. In order to estimate this entity, matching costs among the pixels of the two images is first computed. An energy function is then optimized at which the minimum is the sought depth map. An excellent survey about this topic is found in [1] while up-to-date stereo algorithms can be found at [2].

Once a disparity map is available, it can be used for several purposes, e.g. 3D scene reconstruction, image based rendering (IBR) and 3DTV. The important issue is to have an accurate estimate of the depth map of the scene in order to minimize the errors in the application where it is employed. This is why developing accurate stereo algorithms remain till the moment a hot topic of research, see [2].

Irrespective of the algorithm used to compute it, a disparity map has to be efficiently compressed to either save the hardware storage requirement or before transmitting it over a network as in telepresence applications or in 3DTV. The compression scheme should be designed in a way to simultaneously save the required resources and preserve the quality of the original uncompressed map. A typical way to compress these entities is to use standard image or video compression techniques like JPEG 2000 or MPEG-4. Such schemes process the images in such a way to maintain the visual quality of the result; therefore, they lead to a high amount of error in the reconstructed values of the depth map, see [3–5] for more details.

Motivated by this limitation, we propose in this work an algorithm for depth map compression based on compressed sensing (CS) [6,7]. We first subsample the disparity map in the frequency domain to obtain incoherent measurements. We then formulate a $L_1$-norm conjugate gradient minimization scheme to reconstruct the depth map for the measurements. To ensure the smoothness of the recovered disparity map and preserve the discontinuities at the edges, we incorporate the Total Variation (TV) constraint in the optimization. The results we obtain on several disparity images show that the CS-based scheme leads to an improvement in the quality when compared to other techniques, even at high compression ratio.

The rest of this paper is organized as follows. We give a brief survey on depth map compression techniques in Section 2. We derive in Section 3 the proposed CS based technique for disparity map compression. We test the proposed scheme and compare it to other compression methods in Section 4. Finally, we draw some conclusions and elaborate on some future work in Section 5.

2. RELATED WORK

Depth map compression is an important problem in computer vision. The standard way of tackling it is by applying standard image and video compression techniques like JPEG, JPEG 2000 and MPEG-4. These techniques suffer, however, from several disadvantages if a high compression ratio is desired. JPEG and MPEG-4 lead to blocking artifacts while JPEG 2000 has the effect of blurring the edges or the depth discontinuities [3–5]. This is because these schemes compress a depth image while maintaining the visual quality of the output. This concept is not efficient for the depth maps since the entity is a piecewise smooth surface, i.e. it has discontinuities along the edges of the image while it is smooth otherwise [1]. Applying such concepts might not be harmful when visualizing the depth image but leads to a lot of artifacts and errors in 3D reconstruction and IBR.

These limitations motivated the research to develop more sophisticated depth map compression schemes. In [3], JPEG 2000 was modified to accommodate for region of interest coding and reshaping the dynamic range of the depth map. In [4], coding methods based on adaptive meshing of the depth image were derived. In [9], a depth map was hierarchically decomposed into four regions depending on the locations of the edges. These regions were then merged and fed to an AVC encoder. A context modeling method was derived in [10] to compress the images. A combination of quadtree subdivisions and platelet encoding was developed in [5]. Other methods that can also be applied for depth map compression are based on the combination of non-uniform sampling and adaptive meshing of images as in [11,12] for they take the content of the map into account.

3. PROPOSED SCHEME

Compressed or compressive sensing (CS) is the art of reconstructing a signal from some random measurements that directly condense the
original signal into a compressed representation. The measurements acquired are condensed in the sense that they are very low in number which makes further compression not necessary. CS differs from the traditional data acquisition schemes since it does not follow the sample then compress framework. The bottleneck here lies in the methodology that should be used to reconstruct the original signal from the measurements. CS is relatively a new field of research. It was firstly introduced by Donoho and Candes et al. in [6, 7]. After that, the domain of its applications grew very fast, see [13–15] for some examples. Some good introductory tutorials about CS are presented in [16, 17] while the latest developments are found in [18].

3.1. Problem Formulation

In general, a depth image $X \in \mathbb{R}^{W \times H}$ can be assumed to be compressible or $K$–sparse along its frequency (Fourier) domain. Let $x \in \mathbb{R}^{N \times 1}$ be the vectorized version of the depth map, i.e. $N = W \times H$, and let $\Psi \in \mathbb{R}^{N \times N}$ be the basis matrix of the Fourier transform. The vector $x$ can be written as

$$x = \Psi s,$$

(1)

where $s$ represents the weighting coefficient vector. The goal of our CS based compression scheme is to reconstruct the original depth map $x$ from a measurement vector $y \in \mathbb{R}^{M \times 1}$ where $M \ll N$. This problem can be expressed as

$$y = \Phi x = \Phi \Psi s,$$

(2)

where $\Phi \in \mathbb{R}^{M \times N}$ is the measurement matrix. The matrix $\Phi$ in our problem is a random subsampling matrix of the Fourier transform basis $\Psi$. In order to compress the depth map, we will represent it by the sparse subsampling matrix $\Phi$ and the corresponding measurement vector $y$. Therefore, the first part of our algorithm, i.e. the compressor, is to design a subsampling matrix $\Phi$ which can efficiently compress the original depth map $x$ into the measurement vector $y$ without deteriorating it.

The second part of our algorithm, i.e. the decompressor, is to recover back the original depth map $x$ or equivalently the coefficient vector $s$ in the Fourier basis $\Psi$ from the measurements $y$ based on the theory of compressed sensing. Formulating (2) using a $L1$-norm optimization problem such as

$$\hat{s} = \arg \min_s \|s\|_1 \quad \text{such that} \quad \Phi \Psi s = y,$$

(3)

where $\|\cdot\|_1$ is the $L1$-norm. This equation can also be written as

$$\arg \min_s \left( \|\Phi \Psi s - y\|_2 + \lambda \cdot \|s\|_1 \right).$$

(4)

The right addend of the summation encourages the sparsity of the solution while the left one ensures its consistency with $y$ [6, 7]. The scalar $\lambda$ weighs the contribution the right term in the minimization. However, the properties of the depth map must be additionally taken into account in the minimization of (4). In other words, it is necessary to incorporate the smoothness of the depth map. To do that, we will apply the total variation (TV) constraint. TV was first introduced in [19] for noise removal algorithms and has been successfully applied in estimating the depth maps of stereo sequences [20]. This constraint consists of computing the finite differences of the disparity map to measure the overall amount of variation in the image. The TV constraint ensures the smoothness of the depth map and preserves the discontinuities at the edges at the same time. This is why TV is widely used in image reconstruction problems, see [19–21]. Including this constraint in (4) will encourage the first order derivative to be also sparse. Taking TV into account, the cost function in (4) becomes

$$\arg \min_s \left( \|\Phi \Psi s - y\|_2 + \lambda \cdot \|s\|_1 + \gamma \cdot \text{TV} (\Phi \Psi s) \right),$$

(5)

where $\gamma$ regularizes the weight of TV in the minimization.

What is left to be done is to design a suitable subsampling matrix $\Phi$, i.e. the compressor, to compress the original depth map and a depth map reconstruction algorithm from the compressed signal, i.e. the decompressor, that optimizes (5).

3.2. Designing the Subsampling Matrix

In a general CS problem, the subsampling matrix is chosen as a random matrix. This is done to ensure that the sampling of the signal acts like incoherent aliasing interference. In other words, the random subsampling will enable us to recognize the important coefficients of the sparsifying transform since they remain stronger than the interference caused by the sampling process itself. This will allow us to design a non-linear optimization scheme to recover back the original signal only from the significant coefficients.

Thus, it is important in our algorithm to choose a subsampling matrix $\Phi$ that ensures the above property. To do that, we will apply the Cartesian grid sampling technique of [22]. The method starts by defining a sampling grid equal to the image resolution. The probability density function of the grid is then computed and the subsampling is performed by randomly drawing indices for the density function depending on desired compression ratio. This process is repeated a predefined number of times. Each time, the incoherence is measured via the corresponding point spread function. The sought sampling pattern is then chosen as the one that leads to the best incoherent interference of the Fourier transform.

Once found, the sampling matrix $\Phi$ can be used for any depth map of the specified resolution. This means that it can be preset at the compressor and the decompressor and does not need to be recomputed for another depth map of the same size.

3.3. Depth Map Reconstruction with Conjugate Gradient

To recover the original depth map from the measurements in the CS sense, it is necessary to formulate a non-linear $L1$-norm convex optimization scheme [6, 7]. We will apply here the Fletcher-Reeves non-linear conjugate gradient iterative scheme. The conjugate gradient method computes at each iteration a direction vector $p_k$ for the following update equation

$$s_{k+1} = s_k + \alpha_k p_k,$$

(6)

The constant $k$ is the iteration number while $\alpha_k$ is the line search parameter. This minimization scheme does not require any of the older directions than $p_k$ since the latter is conjugate to each one of them. The new direction of the minimization $p_{k+1}$ is computed from $p_k$ and the residual error of the iteration. Let $f_k$ be the value of the argument in (5) at the $k^{th}$ iteration and $\nabla f_k$ be the corresponding gradient. The new optimization direction is given by

$$p_{k+1} = -\nabla f_{k+1} + \frac{\nabla f_{k+1} \cdot \nabla f_{k+1}^T}{\nabla f_k^T \nabla f_k} p_k.$$

(7)

These two equations are then iterated until convergence is achieved. To ensure that the algorithm does not diverge, the line search parameter $\alpha_k$ must be computed by making it satisfy the strong Wolfe conditions. For more information about the conjugate gradient scheme, the interested reader may refer to [23] for more details.
4. RESULTS AND DISCUSSION

We will perform some tests that consist of evaluating the performance of the proposed CS based compression scheme. We will use as a test data set the ground truth depth maps of the Middlebury test bench [1, 2]. We will compare our scheme with the JPEG and the JPEG 2000 image compression standards. In all the performed tests, the weights $\lambda$ and $\gamma$ of (5) are set to 0.02 and 0.01 respectively.

We will be using two quality measures in the comparisons. The first one is the percentage of the bad disparity values in the depth map while the second is the Peak Signal to Noise Ratio (PSNR). These will be measured while varying the compression ratio at which the ground truth disparity image is compressed. The output of the experiment on the Tsukuba, Cones, Dolls and Art ground truth depth maps is shown in Fig. 1. An interesting phenomenon that can be seen here is that the PSNR is not a good metric to measure the quality of compressed depth maps. The PSNR values obtained from JPEG or JPEG 2000 can be higher than the values obtained with the CS scheme while having higher percentage of error in the depth reconstruction. This can be justified by the well known fact that JPEG and JPEG 2000 are designed to optimize the output for the visual quality of the depth map, i.e. they take into account the human visual psychological model. This leads to a better visualization of the depth map at the risk of obtaining higher error rate of the depth values. As a result, one can conclude from Fig. 1 that PSNR is not a good representative of the quality of a depth map since it can be low even if the corresponding amount of bad depth values is high.

To show the effect of the depth errors visually, we present in Fig. 2 the synthesized right view of the Middlebury images using the original left view and various compressed depth maps. The compression ratio is set to 10. We also provide the output with the ground truth depth map of Middlebury. Note that the occluded regions are left in black since they cannot be observed from the left image. What we can see from the images is that the compression using CS has the least visual effects in the synthesized right images. Moreover, CS has the closest output to that of the ground truth depth maps. This can be clearly seen at the edges of the reconstructed views in Fig. 2.

In all the obtained results, the proposed CS based compression scheme has shown a better performance for the optimization takes the properties of the depth map into account and does not optimize to only maintain the visual quality of the depth map as the others do. To what concerns the speed, the scheme requires 4.6 s on Tsukuba, 10.4 s on Cones and 12.7 s on Art. These results were measured using a 2 GHz 64-bit AMD Opteron which consists of 4 CPUs. The technique was implemented using the C++ programming language.

5. CONCLUSION AND OUTLOOK

We proposed in this paper a new depth map compression technique based on compressed sensing. The method first subsamples the depth map in the frequency domain by choosing a subsampling matrix that ensures the incoherence of the measurements. A non-linear conjugate gradient scheme is then derived to recover the original depth map from the subsampled version. The proposed compression scheme was able to give better results than the standard JPEG and JPEG 2000 image compression algorithms. This is due for the CS based technique does not consider the visual quality in the optimization process. Instead, it takes the properties of the depth map into account by incorporating the TV constraint.

Looking into the future, we want to integrate a suitable coding method with our scheme to obtain higher compression ratio while preserving the quality of the resulting depth map.

6. REFERENCES

Fig. 1. Comparison of the compression algorithms on the Middlebury depth images. Upper row: Percentage of wrong disparity values versus the compression ratio. Lower row: PSNR versus the compression ratio.

Fig. 2. The resulting right frame synthesized using the original left frame of the Middlebury stereo images and various depth maps. From up till down: Tsukuba, Cones and Art. The compression ratio is set to 10.


