VIEW INTERPOLATION CONFIDENCE-AIDED COMPRESSED SENSING OF MULTIVIEW IMAGES

Xing Wang, Jie Liang *

School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

ABSTRACT

In this paper, a hybrid multiview imaging system is considered, where traditional cameras and compressed sensing (CS) cameras are interleavingly placed. To improve the reconstruction quality of the CS cameras, the interpolated image from the two neighboring traditional cameras is used as side information. Different from existing CS-based multiview imaging systems, we incorporate in the CS reconstruction both the frame-level and pixel-level confidences of the interpolated view, based on the knowledge of occluded pixels and holes in it. Simulation results demonstrate the flexibility and superior performance of the proposed framework.

Index Terms— Compressed Sensing, Multiview Images, View Interpolation, Occlusion.

1. INTRODUCTION AND RELATION TO PRIOR WORK

Multiview images are captured by a group of cameras from slightly different locations. Together with new display technologies such as free view-point TV and autostereoscopic displays, an immersive viewing experience can be achieved. However, multiview systems require higher costs for data acquisition, storage and transmission. Fortunately, in most multiview applications, there exist strong correlations between neighboring views. Therefore view-interpolation-based methods can be used to improve the compression efficiency [1, 2]. It can also be used to reduce the acquisition cost. In this paper, a hybrid multiview imaging system is considered, where traditional high-resolution cameras and emerging low-cost compressed sensing (CS) cameras are interleavingly placed. The key idea of the CS theory is that if a signal is sparse in some basis, it can be reconstructed with high quality via simple random sampling at the encoder and ℓ_1-norm optimization at the decoder [3]. Therefore the cost of the CS cameras can be lower than traditional cameras.

A number of CS-based single-view video systems have been proposed that take advantage of the correlations between neighboring frames. In [4], the video frames are divided into non-key and key frames. To reconstruct a non-key frame I_j, the motion-compensated prediction $\widehat{I}_j$ generated from two neighboring key frames is used as the starting point of the Gradient Projection for Sparse Reconstruction (GPSR) algorithm [5]. Several stopping criteria for GPSR are studied. In [6], another distributed compressed video sensing (DISCOS) scheme is proposed, based on the assumption that the difference between I_j and the side information $\widehat{I}_j$ is also sparse in some basis. Therefore, instead of reconstructing the original frame I_j directly, the GPSR is used to reconstruct the residual frame $I_d = I_j - \widehat{I}_j$ from the residual measurement $Y_d = \Phi(I_j - \widehat{I}_j)$. Once the residual frame is estimated, it is added back to the side information to get the final reconstructed frame.

CS-based multiview imaging systems have also been studied. In [7], the side information for CS reconstruction is generated from neighboring views by disparity estimation instead of motion estimation. The CS recovery is based on projection-based reconstruction, a specific instance of a projected Landweber algorithm. In [8], a different approach to exploit the interpolated side information from neighboring cameras is developed, based on the belief propagation-based compressive sensing framework (BPCS) [9]. The method in [8] also uses the side information as the starting point for the belief propagation. One limitation of BPCS is that it assumes that different transform coefficients have the same distribution, which is not the case in natural images and videos.

However, existing multiview imaging systems in [7, 8] have not fully exploited all information in view interpolation. First, it is known that view interpolation quality is highly dependent on the scene composition. Therefore, based on the overall frame-level confidence of the interpolated image provided by the view interpolation algorithm, we should have a mechanism to adjust the influence of the view interpolation result on the CS reconstruction.

Secondly, many view interpolation algorithms also provide confidence information at pixel level [2], in terms of the number of matching points a pixel of the interpolated view can have in the two neighboring views. Usually, pixels with two matching points have higher reconstruction quality. Pixels with only one matching point are occluded in one neighboring

*This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under grant RGPIN312262. Emails: {xingw, jiel}@sfu.ca.

view, thereby having lower interpolation quality. For pixels without any correspondence in the neighboring views (corresponding to holes in the initial interpolated image), various inpainting methods have to be used to estimate their values. Therefore these pixels generally have the lowest confidence.

If these issues are not addressed properly, existing view interpolation-aided multiview CS reconstruction methods could perform even worse than direct CS reconstruction. In this paper, we propose a modified GPSR algorithm by adding another term to the objective function. The term measures the squared error between the CS-based and view-interpolation-based reconstructions. The weighting parameters of this term are determined by both the frame-level and pixel-level confidences of the view interpolation result. We show that the modified method can still be converted to the GPSR framework. Simulation results demonstrate that the framework is very flexible and can outperform existing methods.

2. BACKGROUND OF CS AND GPSR

Assume x is a length-N, real-valued signal whose decomposition θ in an orthonormal basis Ψ has K significant coefficients (K sparse), *i.e.*, $x = \Psi\theta$. In compressed sensing, the measurement is obtained by a simple linear operator Φ of size $M \times N$, where $M \ll N$. That is, $y = \Phi x = \Phi\Psi\theta = A\theta$ [3]. Typical measurement matrices used in CS include Gaussian, sub-Gaussian, or Bernoulli matrices.

The problem of reconstructing x or θ from y is underdetermined. However, since θ is sparse, the ℓ_1 optimization can be used, for example,

$$\arg\min_{\theta} \|\theta\|_1, \quad \text{s.t. } y = A\,\theta. \tag{1}$$

The problem can be efficiently solved via linear programming. However, for large-scale applications, the speed of the optimization algorithms can be very slow. Recently, a fast Gradient Projection for Sparse Representation (GPSR) algorithm has been developed [5], which starts with the following unconstrained convex optimization problem

$$\arg\min_{\theta} \left(\frac{1}{2} \|y - A\theta\|_2^2 + \tau \|\theta\|_1 \right), \tag{2}$$

where τ is a weighting parameter that enforces the sparsity constraint.

To solve this, it first decomposes θ into its positive and negative parts.

$$\theta = u - v, u \geq 0, v \geq 0. \tag{3}$$

The problem can then be converted to the following bound-constrained quadratic programming (BCQP) formulation of basis pursuit or similar problems [3].

$$\begin{aligned} &\min_{z} c^T z + \tfrac{1}{2} z^T B z \equiv F(z), \\ &\text{s. t. } z \geq 0, \end{aligned} \tag{4}$$

where

$$\begin{aligned} &z = \begin{bmatrix} u \\ v \end{bmatrix}, \ b = A^T y, \ c = \tau 1_{2N} + \begin{bmatrix} -b \\ b \end{bmatrix}, \\ &B = \begin{bmatrix} A^T A & -A^T A \\ -A^T A & A^T A \end{bmatrix}. \end{aligned} \tag{5}$$

The solution to (4) is equal to the solution of (2) if the free parameter τ is much less than 1.

It is shown in [5] that good solutions can be obtained very fast by using gradient projection, special line search and termination techniques, making the GPSR method very attractive.

3. GENERALIZED GPSR WITH VIEW INTERPOLATION CONFIDENCE

In this section, we propose a generalized optimization framework to consider the occlusions and holes in the interpolated image. We then show that the framework can be converted into the standard BCQP format, which can be efficiently solved by the GPSR algorithm.

3.1. Generalized Optimization Framework

Our goal is to reconstruct the middle image I_j from its linear CS measurements y, with the help of the interpolated middle image $\widehat{I}_j$ generated from the left and right reference images I_{j-1}, I_{j+1} (given by conventional cameras). Due to the strong correlation between images in multi-view image system, the final reconstructed image should be generally close to the interpolated image. However, the quality of the interpolated image $\widehat{I}_j$ is affected by the number of occlusion pixels and the size of the holes in it. Hence, if we reconstruct the difference image between I_j and $\widehat{I}_j$ and add it back to $\widehat{I}_j$ to get the reconstructed image, the performance could be even worse than directly reconstructing the middle image from its CS measurement, because the sparsity of the difference image could be larger than the sparsity of the original middle image in this case.

To resolve this potential issue, we propose the following generalized optimization framework.

$$\arg\min_{\theta} \left(\frac{1}{2} \|y - A\theta\|_2^2 + \tau \|\theta\|_1 + \frac{\mu}{2} \sum_{i=1}^{N} w_i \Big(I_{i,j} - \widehat{I}_{i,j} \Big)^2 \right), \tag{6}$$

where θ is the sparse representation of I_j in basis Ψ, *i.e.*, $I_j = \Psi\theta$. The last squared-error term is new compared to the original GPSR in (2). $I_{i,j}$ and $\widehat{I}_{i,j}$ denote the i-th pixel of the target image I_j and the interpolated image $\widehat{I}_j$, respectively. μ is a weighting parameter that is determined by the overall frame-level confidence of the view-interpolation algorithm, and w_i is the weighting parameter for the i-th pixel, which is determined by the pixel-level view interpolation confidence.

A larger value of μ can be used if the overall view interpolation has higher quality. In this case, the view interpolation is more trustworthy. On the other hand, μ should be smaller if there are many occlusion pixels and holes in the view interpolation; hence the CS reconstruction should rely more on the linear measurement from the CS camera.

Similarly, the pixel-level weighting parameter w_i should be larger when a pixel in the middle view has two point correspondences in the neighboring views. A smaller w_i should be used when there is only one point correspondence, *i.e.*, the pixel is occluded in one view. Finally, the smallest w_i should be used when no point correspondence can be found, as the pixel is in a hole in the initial view interpolation. The occluded pixels and holes usually occur near the edges of objects in an image.

The impacts of μ and w_i will be studied in Sec. 4.

3.2. Conversion to the Standard BCQP Format

Let $\widehat{\theta}$ be the sparse representation of the interpolated image $\widehat{I}_j$ in basis Ψ, ψ_i the i-th row of Ψ, and $R_i = \psi_i^T \psi_i$, which is a symmetric matrix. Each squared error in the last term of (6) can be written as

$$(I_{i,j} - \widehat{I}_{i,j})^2 = (\theta - \widehat{\theta})^T R_i (\theta - \widehat{\theta}) \ . \tag{7}$$

As in (3), we split θ and $\widehat{\theta}$ into their positive and negative parts. The generalized framework in (6) can thus be converted to the standard BCQP format in (4), with the following definitions:

$$\begin{aligned}
z &= \begin{bmatrix} u \\ v \end{bmatrix}, \quad b = A^T y + \mu \sum_{i=1}^{N} w_i R_i (\widehat{u} - \widehat{v}), \\
c &= \tau 1_{2N} + \begin{bmatrix} -b \\ b \end{bmatrix}, \\
B &= \begin{bmatrix} A^T A + \mu \sum_{i=1}^{N} w_i R_i & -(A^T A + \mu \sum_{i=1}^{N} w_i R_i) \\ -(A^T A + \mu \sum_{i=1}^{N} w_i R_i) & A^T A + \mu \sum_{i=1}^{N} w_i R_i \end{bmatrix}.
\end{aligned} \tag{8}$$

The GPSR algorithm can then be used to solve this BCQP problem.

4. SIMULATION RESULTS

In this section, we present some simulation results to compare our proposed algorithm with other GPSR-based algorithms. The orthonormal basis in the CS is chosen as the DCT. In the image acquisition step of compressed sensing, the scrambled block Hadamard ensemble (SBHE) method proposed in [10] is used. The size B and free parameter τ are chosen according to [10]. The view interpolation method in [1] is used.

In the following experiments, *View-Interp* represents the view interpolation result given by the method in [1], which will be included in the figures as a reference. *Direct-GPSR* refers to a method similar to the scheme in [4], where the view interpolation result $\widehat{\theta}$ is directly used as the initial value of GPSR reconstruction. *Diff-GPSR* is the generalization of [6] to multiview image systems, where the GPSR method is used to recover the residual frame, which is then added back to the interpolated image to get the final reconstruction.

VIC-GPSR denotes the proposed view interpolation confidence (VIC)-aided GPSR method. Its initial value is also chosen as $\widehat{\theta}$. The frame-level weighting parameter μ is selected for each multiview image data set, as will be described below. The pixel-level weighting parameter w_i is chosen to be 1, 0, and 0 respectively, if a target pixel has two, one or zero point correspondence in view interpolation. That is, in the interpolated image, we only trust the pixels with two point correspondences when evaluating the CS reconstruction. Though from the Eq.8, it seems that we need to operate on the columns of basis Ψ, which highly increases the complexity, in fact, it can operate on the whole matrix in which case we can use fast algorithms to do the transformation.

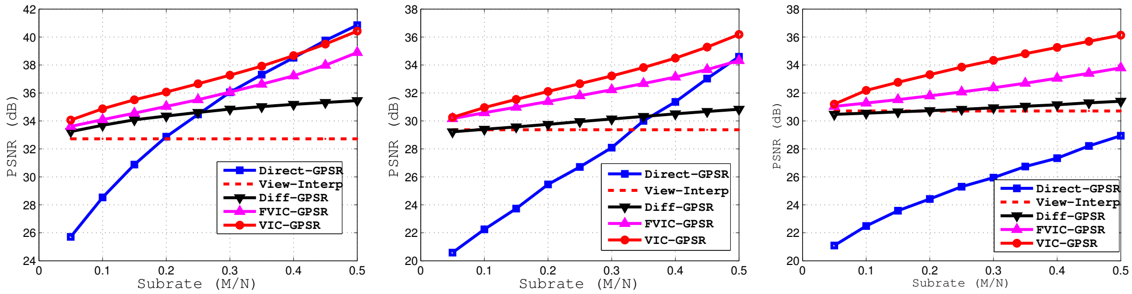
FVIC-GPSR is another special case of the proposed method, where all w_i's are fixed to be 1. In this case, the pixel-level view interpolation confidence is not exploited, and only the frame-level weighting parameter μ is in effect.

In the following, the multiview video datasets Akko & Kayo, Christmas and Teddy are used, with frame size of 640×480, 640×480, and 448×352, respectively. The first and third views of each dataset are assumed to be given by traditional cameras, and the second view is assumed to be sampled by a CS camera and reconstructed by different CS algorithms. Only the first frame of each view is tested.

Fig. 1 (a) shows the reconstruction PSNR versus CS sampling subrate M/N of different methods with the multiview image dataset Akko & Kayo. The view interpolation result shows that the number of target pixels with two, one and zero point correspondences is 285879, 20317, and 1004, respectively. The weight parameter μ is chosen to be 1.

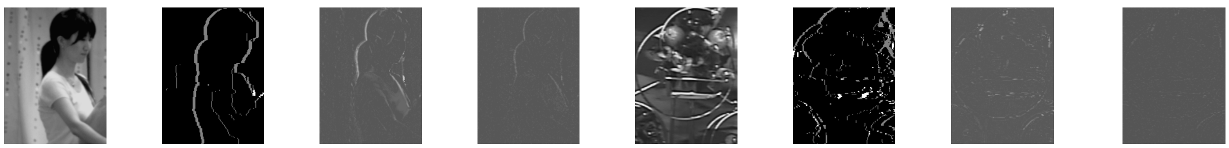
Some observations can be made from Fig. 1 (a). First, at low subrate, *Direct-GPSR* is much worse than other methods. As the number of samples M increases, *Direct-GPSR* can get close to and eventually outperform other methods, including our proposed *VIC-GPSR*. The reason is that the parameter μ is fixed to 1, which essentially gives the same weight to the first and the last term in Eq. (6).

Secondly, the proposed *VIC-GPSR* and *FVIC-GPSR*, as well as *Diff-GPSR* can always have better results than the interpolated view. Our methods also always achieve better results than *Diff-GPSR*, and the gain increases with the subrate (more than 3 dB when the subrate is greater than 0.3), which shows the power of considering the view interpolation confi-



(a) (b) (c)

Fig. 1. PSNRs versus sampling subrate of different methods. (a) Akko & Kayo. (b) Christmas. (c)Teddy.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Original image (a), VIC maps (b), reconstruction errors of Diff-GPSR (c) and VIC-GPSR (d) for Akko & Kayo, as well as original image (e), VIC maps (f), reconstruction errors of Diff-GPSR (g) and VIC-GPSR (h) for Christmas.

dence information.

Third, *VIC-GPSR* has better performance than *FVIC-GPSR*, thanks to the contribution of the pixel-level confidence information. The gain also increases with the subrate.

Fig. 1 (b) are the results using the dataset Christmas. The view interpolation result shows that the number of target pixels with two, one and zero point correspondences is 272428, 31029, and 3743, respectively. This means that the view interpolation of this dataset is not as good as that in Akko & Kayo, as indicated by the PSNRs of the view interpolation method in Fig. 1 (a) and Fig. 1 (b). In our methods, μ is set to be 1.

Fig. 1 (b) shows that *Direct-GPSR* is worse than *View-Interp* when subrate is less than about 0.33. This verifies that if the view interpolation does not have good quality, directly using it as the initial value of GPSR could lead to even worse result than the interpolated view. Our methods and *Diff-GPSR* can still outperform the view interpolation. Note that *Diff-GPSR* only has limited gain over the view interpolation method, and the gain of our methods over *Diff-GPSR* can be more than 4 dB when the subrate is greater than 0.3.

Fig. 1 (c) shows the results with the Teddy dataset. The number of target pixels with two, one and zero point correspondences is 141946, 15289, and 461, respectively. Therefore, we choose the weighting parameter μ to be 5.

It can be seen from Fig. 1 (c) that *Diff-GPSR* achieves almost the same result as *View-Interp*, sometimes even worse when the subrate is low, because the subrate is not enough to recover the difference image accurately, and fails to capture the edges in the middle image. Our proposed algorithm always achieves the best performance.

Fig. 2 shows portions of the original image, the view interpolation confidence (VIC) maps, the final reconstrutcion errors of Diff-GPSR and the proposed VIC-GPSR for two images. The locations of occluded pixels and initial holes are represented by gray and white pixels in the VIC maps. It can be seen that these pixels are near the sharp edges of the images. The proposed method clearly has much less reconstruction errors near the sharp edges. This verifies that our method can avoid the adverse impact of the occlusions and holes in view interpolation.

5. CONCLUSION AND FUTURE WORK

In this paper, we consider view interpolation-aided compressed sensing of multiview images. Different from existing methods, we exploit the knowledge of occlusions and holes in the interpolated view when performing the CS reconstruction, by assigning more weights to the view interpolation result when its quality is satisfactory, and vice versa. Experimental results show that our method can outperform existing CS-based multiview image systems.

The proposed scheme opens up some new topics for future research. For example, how to automatically determine the weighting parameters μ and w_i, and how to generalize it to multiview video systems. Another possible way to further improve the performance is to apply the iterative method in [11].

6. REFERENCES

[1] X. Xiu, D. Pang, and J. Liang, "Rectification-based view interpolation and extrapolation for multiview video coding," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 21, no. 6, pp. 693–707, Jun. 2011.

[2] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Cambridge Univ. Press*, 2003.

[3] D. Donoho, "Compressed sensing," *IEEE Transaction on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 1169–1172.

[5] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, pp. 586–597, Dec. 2007.

[6] T. Do, Y. Chen, D. T. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *IEEE International Conference on Image Processing*, 2009, pp. 1393–1396.

[7] M. Trocan, T. Maugey, J. E. Fowler, and B. Pesquet-Popescu, "Disparity-compensation compressed-sensing reconstruction for multiview images," in *IEEE International Conference on Multimedia and Expo*, 2010, pp. 1225–1228.

[8] P. Beigi, X. Xiu, and J. Liang, "Compressive sensing based multiview image coding with belief propagation," in *Proc. Asilomar Conference on Signals, Systems, and Computers*, 2010, pp. 430–433.

[9] R. Baron, S. Sarvoham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Proc.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[10] L. Gan, T. T. Do, and T. D. Tran, "Fast compressive imaging using scrambled block hadamard ensemble," in *European Signal Processing Conf. (EUSIPCO)*, 2008.

[11] M. Trocan, T. Maugey, Eric. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Compressed sensing of multiview images using disparity compensation," in *IEEE International Conference on Image Processing*, 2010, pp. 3345–3348.