OPTICAL FLOW FOR COMPRESSIVE SENSING VIDEO RECONSTRUCTION

H. Braun, P. Turaga, C. Tepedelenlioglu, A. Spanias

School of ECEE, SenSIP Center and Industry Consortium,
Arizona State University
{hcbraun, pturaga, cihan, spanias}@asu.edu

ABSTRACT

Although considerable effort has been devoted to the problem of reconstructing compressively sensed video, no existing algorithm achieves results comparable to commonly available video compression methods such as H.264. One possible avenue for improving compressively sensed video reconstruction is the use of optical flow information. Current efforts reported in the literature have not fully utilized optical flow information, instead focusing on limited cases such as stationary backgrounds with sparse foreground motion. In this paper, a reconstruction method is presented which fully utilizes optical flow information to increase the quality of reconstruction. The special cases of known image motion and constant global image motion are presented, and the performance of the algorithm on existing datasets is evaluated.

***Index Terms*—** Image Reconstruction, Compressive Sensing, Optical Flow, Motion Estimation

1. INTRODUCTION

Reconstruction of compressively sensed video remains an open problem. Progress has been made in reconstruction of static images [1], but no currently published algorithm fully utilizes prior knowledge of frame-to-frame correlations and compressibility to improve reconstruction. Previous efforts have made restrictive assumptions such as a static background [2, 3] or constant optical flow across the entire image [4]. Clearly, a reconstruction algorithm which fully utilizes dense optical flow information to perform multi-frame video reconstruction would be a valuable tool and would bring compressive sensing video hardware closer to commercialization.

In this paper, a method of reconstructing compressively sensed video frames in the presence of known optical flow is presented in Section 2.1 and applied to an existing dataset in Section 3.1. This approach is extended to the case of unknown but constant motion in Section 2.2, with results presented in Section 3.2. In both cases, significant performance improvements are achieved by incorporating optical flow information in reconstruction. It is hoped that this work represents a step toward fast and reliable estimation of dense optical flow and wide adoption of compressive sensing video technology.

This work is funded by Raytheon Missile Systems and supported in part by the SenSIP site of the NSF NCSS I/UCRC under award SO322.

1.1. Problem Statement

The problem of reconstructing a compressively sensed video is briefly described in this section, and the notation used throughout the paper is introduced. Two image vectors, $\mathbf{x}_1$ and $\mathbf{x}_2$, are known to be compressible. That is, they are known to be sparse in some basis with inverse transform operator B. These images are measured with sensing matrix M, resulting in measurements $\mathbf{y}_1$ and $\mathbf{y}_2$. That is,

$$\mathbf{y}_1 = \mathrm{M}\mathbf{x}_1 = \mathrm{MB}\boldsymbol{\theta}_1 \tag{1}$$

$$\mathbf{y}_2 = \mathrm{M}\mathbf{x}_2 = \mathrm{MB}\boldsymbol{\theta}_2, \tag{2}$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are k-sparse vectors forming compressed representations of $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. M and B may be chosen in one of several ways. Some of the most popular methods are an IID Gaussian matrix, a random orthoprojector, or a noiselet-based pseudorandom matrix. M is a $n \times m$ fat matrix, where n is the number of compressive measurements to be taken and m is the number of pixels in an image. For natural images, wavelet transforms and discrete cosine transforms provide high sparsity and are good candidates for B. Throughout this paper, an IID Gaussian sensing matrix and Daubechies-4 wavelet basis will be used.

Since $n < m$, (1) and (2) are under-determined systems and cannot be explicitly solved for $\mathbf{x}_1$ and $\mathbf{x}_2$. However, it is well-known that $\boldsymbol{\theta}_1$ can be recovered with high probability from $\mathbf{y}_1$ by solving the convex optimization problem

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \; \|\mathrm{MB}\boldsymbol{\theta} - \mathbf{y}_1\|_2 + \tau \|\boldsymbol{\theta}\|_1 . \tag{3}$$

Many solvers have been developed for problems in the form of (3), including GPSR [5], and SPGl1 [6, 7]. In addition, several optimization methods exist to approximately solve a non-convex ℓ_0 minimization form of the problem[8, 9].

In the case of compressively sensed video, it is also known that $\mathbf{x}_1$ and $\mathbf{x}_2$ are related by some linear optical flow operator

$\mathrm{F}_{\mathbf{u},\mathbf{v}}$, where $\mathbf{u}$ and $\mathbf{v}$ are optical flow vectors in the x- and y-directions, respectively. That is,

$$\mathbf{x}_2 = \mathrm{F}_{\mathbf{u},\mathbf{v}}\mathbf{x}_1 + \boldsymbol{\eta}, \tag{4}$$

where $\boldsymbol{\eta}$ is an error term representing the part of $\mathbf{x}_2$ which is independent of $\mathbf{x}_1$. In natural images, $\boldsymbol{\eta}$ is sparse, a fact which is exploited by the majority of video codecs used today.

1.2. Previous Work

In the field of video reconstruction, batch algorithms have been proposed which recover multiple frames simultaneously without taking advantage of correlations between frames[10]. There are also several examples in the literature of attempts to use multi-frame information for compressively sensed video reconstruction. However, all make restrictive assumptions on the nature of the video sequence being reconstructed. In this section, several existing algorithms are described, along with their benefits and problems. Table 1.2 summarizes the characteristics of the algorithms described in this section.

In [4], global optical flow is estimated from compressive measurements of two video frames and then used to reconstruct both frames. While useful, this method requires the two input frames to be pure translations of one another. When optical flow is not constant across the entire image or is large (on the order of several pixels), the algorithm fails. This method also requires the use of a highly unorthodox sensing scheme in which compressively sensed pixels are compactly supported on the image. A sensing matrix of this type is likely to have undesirable properties.

In [11], dense optical flow is estimated from compressive measurements without performing any reconstruction. When the measurement rate is high, good estimates of optical flow are produced. However, an unconventional sensing scheme is again used, in which each measurement is compactly supported in one of the image's spatial dimensions. In fact, each compressive measurement is restricted to a single row of the uncompressed scene.

In [2], a quasi-static background with small moving targets is assumed. This assumption allows a video sequence to be represented as the sum of a low rank matrix, representing the background, and a sparse matrix, representing moving objects in the foreground. The SPARCS algorithm, a greedy pursuit derived from the CoSAMP algorithm [9], is then used to recover the video sequence. Although the SPARCS algorithm performs well when the assumptions of static background and small motion are satisfied, it quickly fails when the background is not static or when targets are large. In addition, a relatively large number of frames (typically hundreds) are needed to accurately estimate the background. This is most likely a consequence of the algorithm's failure to take advantage of the known compressibility of the background image when performing reconstruction.

In [3] a static background with small targets is again assumed. In this case, however, the difference between two images is reconstructed. This is possible because as long as targets are small, the difference image is known to be spatially sparse. This algorithm achieves good results as long as all assumptions are met.

The CS-MUVI algorithm [12] most nearly achieves the goal of fully utilizing known information, including both optical flow and compressibility of individual frames. In this algorithm, a special sensing matrix M is used which creates a well-conditioned matrix when combined with an upsampling operator U. The combined matrix MU is then used to reconstruct a low-resolution version of the video using least-squares estimation. Optical flow is then estimated using this low-resolution video and the resulting optical flow is used to improve reconstruction of the compressively sensed high-resolution video. This method, while highly effective, fails to reconstruct small objects since these objects are not visible in the low-resolution version of the video.

2. ALGORITHM DESCRIPTION

2.1. Reconstruction with Known Optical Flow

In the special case where optical flow vectors $\mathbf{u}$ and $\mathbf{v}$ are known *a priori*, reconstruction is relatively straightforward and the benefits of incorporating optical flow in a reconstruction algorithm are clear. The case of reconstruction from two frames using forward prediction is presented here, although this method is easily extended to include larger numbers of frames and backward prediction. In this case, the underdetermined system to be solved may be jointly given in matrix form by combining (1), (2), and (4), and is described by

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{MB} & 0 \\ \mathrm{MF}_{\mathbf{u},\mathbf{v}}\mathrm{B} & \mathrm{M} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{bmatrix}. \tag{5}$$

$\boldsymbol{\theta}_1$ can then be recovered from $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{u}$, and $\mathbf{v}$ by solving the convex optimization problem

$$\underset{\boldsymbol{\theta},\boldsymbol{\eta}}{\text{minimize}} \left\| \begin{bmatrix} \mathrm{MB} & 0 \\ \mathrm{MF}_{\mathbf{u},\mathbf{v}}\mathrm{B} & \mathrm{M} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{bmatrix} - \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \right\|_2 + \tau \left\| \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\eta} \end{bmatrix} \right\|_1. \tag{6}$$

Equation (6) differs from the optimization problem used in [12] in several important ways. First, optical flow appears as part of the objective function rather than as a constraint, allowing only $\boldsymbol{\theta}_1$ to be estimated and reducing the size of the estimation problem. Second, the inclusion of $\|\boldsymbol{\eta}\|_1$ in the objective function creates sparsity in the optical flow error η. This should allow the more accurate reconstruction of small objects and edges in the final image. The algorithm also differs from that of [4] in that no additional special properties of the sensing matrix are required to perform motion estimation.

Table 1. Comparison of existing video reconstruction algorithms incorporating temporal information.

Name	Video sequence restrictions	Sensing restrictions	Comment
Jacobs et al. [4]	Constant motion	Compactly supported sensors	
Thirumalai & Prossard [11]	None (dense optical flow)	Restricted to one row of image	High sensing rate
SparCS [2]	Static background, small targets	None	Batch algorithm
Background Subtraction [3]	Static background, small targets	None	
CS-MUVI [12]	None (dense flow)	Hadamard-derived sensing matrix	
Proposed Algorithm	Constant or known motion	None	

The case of $\boldsymbol{\eta} \approx \mathbf{0}$ will also be examined; this is a reasonable assumption for several useful applications of compressive sensing. If $\boldsymbol{\eta} \approx \mathbf{0}$, (6) becomes

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \quad \left\| \begin{bmatrix} \mathrm{MB} \\ \mathrm{MF}_{\mathbf{u},\mathbf{v}}\mathrm{B} \end{bmatrix} \boldsymbol{\theta} - \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \right\|_2 + \tau \left\|\boldsymbol{\theta}\right\|_1. \quad (7)$$

2.2. Reconstruction with Constant but Unknown Optical Flow

An interesting special case arises when optical flow is known to be constant across an entire image; that is, when $\mathbf{u}$ and $\mathbf{v}$ are constant vectors. This is generally the case for aerial surveillance video, an important potential application of compressive sensing video technology. In this case, only a single frame must be reconstructed in order to estimate flow. This is possible because of the ability to perform correlations in the compressive domain [13].

Let $\hat{\mathbf{x}}_1$ be an estimate of $\mathbf{x}_1$ acquired by solving (3). $\mathbf{u} = u\mathbf{1}$ and $\mathbf{v} = v\mathbf{1}$ may then be estimated maximizing the estimated correlation with a matched filter:

$$\underset{u,v}{\text{maximize}} \ \mathbf{y}_2^\top \left(\mathrm{MM}^\top\right)^{-1} \mathrm{MF}_{u\mathbf{1},v\mathbf{1}}\hat{\mathbf{x}}_1. \quad (8)$$

Once $\mathbf{u}$ and $\mathbf{v}$ have been determined, two-frame reconstruction may be carried out as in (7) or (6). Although (8) appears straightforward, several practical considerations arise when attempting to apply it. Equation (8) is, in general, a non-convex problem and is relatively expensive to compute. However, if u and v are quantized, an exhaustive search is often feasible, for instance if the image size is small or the motion vector magnitude $\|[u, v]\|_2$ is bounded. Several efficient search algorithms also exist to approximately determine optical flow; Zhu and Ma's diamond search algorithm [14] was used in this work . The correlation is also heavily affected by the way the flow operator handles the edges of the image. In order to avoid this concern, $\hat{\mathbf{x}}_1$ is first windowed before applying the optical flow transform. It was also empirically determined that reconstruction was superior when optical flow was non-zero, even if the ground truth optical flow was zero. Because of this, u and v are forced to take non-zero values: if a zero value is found to be optimal, the next-best value is chosen instead.

3. RESULTS

3.1. Known Optical Flow

The algorithm in Section 2.1 was tested using the Middlebury optical flow dataset's "Venus" image, a stereo image for which known ground truth motion vectors are provided [15]. The image was preprocessed by shrinking by a factor of 2 and cropping to size 128×128. Sensing was performed using an IID Gaussian sensing matrix using $n = 4915$ measurements, for a sensing rate of 0.3. Figure 3.1 shows the results of reconstruction of a single frame, along with simultaneous reconstruction of both frames using known optical flow information. The PSNR was increased from 21.12 dB to 23.80 dB by the use of optical flow information. An improvement in the perceptual quality of the reconstruction is also visible. Figure 3.1 shows PSNR vs. number of sensors for single-frame and 2-frame reconstruction. The benefits are clear, particularly at low sensing rates.

3.2. Constant Optical Flow

The problem of reconstruction under constant but unknown optical flow was evaluated using the PETS2000 dataset. A segment of the video with no activity was selected and used to generate a pair of 128×128 test frames, represented in vector form as $\mathbf{x}_1$ and $\mathbf{x}_2$, in which $\mathbf{x}_2$ experienced a small shift relative to $\mathbf{x}_1$. $\mathbf{y}_1$ and $\mathbf{y}_2$ were generated using $n = 1024$ sensors, and the algorithm described in Section 2.2 was used to reconstruct the image. Figure 3.2 compares single-frame and two-frame optical flow based reconstruction algorithms. The two-frame reconstruction is clearly perceptually superior and has a PSNR of 27.21 dB, a 2.81 dB improvement over the single-frame reconstruction. Reconstruction was performed both with a term $\boldsymbol{\eta}$ for imperfect optical flow (Equation 6) and without it (Equation 7). The $\boldsymbol{\eta}$ term was found to be necessary in order to achieve improved PSNR relative to the single-frame method; only results from this method are presented.

4. CONCLUSION

Optical flow is a key element of modern video codecs, but has not been fully used in reconstruction algorithms for compressively sensed video sequences. This is mainly due to the

(a) ground truth

(b) single-frame

(c) two-frame

Fig. 1. Reconstruction with known optical flow for 128×128 image with $n = 3481$ measurements (sensing rate = 0.212).

difficulty of estimating optical flow without an existing high-quality reconstruction of the image. However, once optical flow is known, clear benefits are available from utilizing optical flow information. This was shown in Section 3.1, where the problem of reconstruction under known optical flow was presented as a convex optimization.

A potential work-around to the problem of optical flow estimation without a high-quality image was described in Section 2.2. In the special case considered, optical flow is known to be constant across the entire image.This allows optical flow in subsequent frames to be estimated without reconstruction from a single reconstructed frame. The authors believe that this method is likely to be applicable to the case of small moving targets in a moving frame, as the optical flow error parameter $\boldsymbol{\eta}$ should be able to accommodate them. However, this claim is speculative and is beyond the scope of this paper.

The authors hope that this work will serve to advance the development of compressive sensing video cameras by integrating existing ideas of optical flow based reconstruction from conventional video codecs. A robust reconstruction algorithm with high performance could dramatically reduce the cost and data rate requirements of video sensing systems at a wide range of light wavelengths.

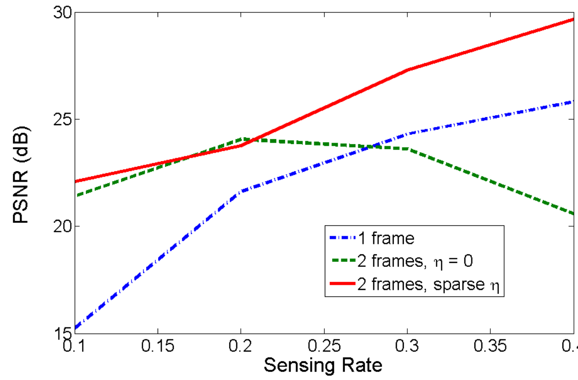

Fig. 2. PSNR of reconstructed vs. original signal as a function of number of compressive measurements.

(a) ground truth

(b) single-frame

(c) two-frame

Fig. 3. Comparison of single-frame and two-frame reconstruction for constant optical flow.

5. REFERENCES

[1] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "An architecture for compressive imaging," in *Image Processing, 2006 IEEE International Conference on*. IEEE, 2006, pp. 1273–1276.

[2] A. Waters, A. Sankaranarayanan, and R. Baraniuk, "SpaRCS: Recovering low-rank and sparse matrices from compressive measurements," *Proc. Adv. in Neural Processing Systems (NIPS), Granada, Spain*, 2011.

[3] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, R. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," *Computer Vision–ECCV 2008*, pp. 155–168, 2008.

[4] N. Jacobs, S. Schuh, and R. Pless, "Compressive sensing and differential image-motion estimation," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 718–721.

[5] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, 2007.

[6] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[7] ——, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008. [Online]. Available: http://link.aip.org/link/?SCE/31/890

[8] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[9] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[10] R. Marcia and R. Willett, "Compressive coded aperture video reconstruction," in *Proc. European Signal Processing Conf.(EUSIPCO)*, 2008.

[11] V. Thirumalai and P. Frossard, "Correlation estimation from compressed images," *ArXiv*, 2011.

[12] A. Sankaranarayanan, C. Studer, and R. Baraniuk, "CS-MUVI: Video compressive sensing for spatial-multiplexing cameras," in *Computational Photography (ICCP), 2012 IEEE International Conference on*, april 2012, pp. 1 –10.

[13] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "The compressive matched filter," Rice University, Tech. Rep., 2006.

[14] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block matching motion estimation," in *Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on*, vol. 1, sep 1997, pp. 292 –296 vol.1.

[15] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.