COMPRESSIVE SENSING AND DIFFERENTIAL IMAGE-MOTION ESTIMATION

Nathan Jacobs, Stephen Schuh, Robert Pless

Washington University in St. Louis, MO, 63130, USA
(jacobsn|schuhs|pless)@cse.wustl.edu

ABSTRACT

Compressive-sensing cameras are an important new class of sensors that have different design constraints than standard cameras. Surprisingly, little work has explored the relationship between compressive-sensing measurements and differential image motion. We show that, given modest constraints on the measurements and image motions, we can omit the computationally expensive compressive-sensing reconstruction step and obtain more accurate motion estimates with significantly less computation time. We also formulate a compressive-sensing reconstruction problem that incorporates known image motion and show that this method outperforms the state-of-the-art in compressive-sensing video reconstruction.

***Index Terms*—** Image sampling, Image motion analysis, Image reconstruction

1. INTRODUCTION

Novel camera designs are sampling the visual environment in ways significantly different from standard pinhole cameras. Recently, the interest in compressive sensing has led to camera system designs based on measurements that sum the light captured over arbitrarily shaped and highly overlapping parts of the visual field. A camera which simultaneously captures many such measurements, which we call an integral-pixel camera, is a useful addition to the toolbox of a vision system designer because it offers new trade-offs in terms of constraints on sensor hardware design.

These integral-pixel cameras are currently being explored within the context of compressive sensing (which implies a specific constraint on the sampling of different pixels), with the goal of reconstructing high resolution representations of a scene with significantly fewer measurements. Successes in this domain include the ability to make images in wavelengths for which it is expensive to make each light sensitive element [1], or to approximate large parts of the 4D light-field incident on a camera [2]. However, there has not been a systematic study of whether the integral-pixels used for compressive sensing are also good for tasks other than image reconstruction. In this paper we explore the use of integral-pixels in scenarios with image motion, both for the estimation of image motion and the reconstruction of translating images.

The major contribution of this paper is an exposition of the interaction between image motion and integral-pixel measurements. Our first contribution is an illustration of how to estimate translational motion for integral-pixel cameras. Second, we explicitly consider the relationship of image motion and compressive-sensing models, and show that in the context of reconstructing translating video, using known motion and the optic-flow constraints offers better reconstructions than other compressive-sensing video reconstruction methods that have been proposed. This suggests that future sensor designs may explicitly include integral-pixels optimized for estimating motions as well as patterns optimized for reconstructing the image.

1.1. Related Work

Recent research in camera design has explored alternatives to traditional pixels that integrate the visual world using small regularly-spaced non-overlapping regions. This includes capturing an image using time-coded shutters [3], coded combinations of locations [4, 1, 5], and locations and wavelengths simultaneously [6]. Other work includes the use of coded apertures [7, 8].

Usually a reconstruction algorithm is used to generate an approximation of the image, hyper-spectral data cube, or light-field as if it was sampled by a regular pixel grid. Although two papers very recently consider the problem of independent motion detection [9], and reconstructing multiple frames of a video [10], both focus on reconstructing the difference images or the video frames rather than estimating the frame-to-frame motion. We show that in some cases it is possible, and often advantageous, to estimate image motion without the computationally-expensive reconstruction step.

2. DIRECT MOTION ESTIMATION WITH INTEGRAL PIXELS

The integral-pixel imaging model considers a generalized version of the sampling scheme of a traditional camera. Let $I(r)$ be the response of a standard camera at particular pixel r. Within the more general integral-pixel imaging model, we consider a camera whose pixels capture a weighted integral of the image. That is, each pixel p_i measures an intensity equal to:

$$\hat{I}(p_i) = \iint I(r) w_i(r) dr, \tag{1}$$

where $w_i(r)$ is a weighting function[1] describing how pixel i samples the image. Note that for traditional cameras $w_i(r)$ is non-zero in a small region.

Suppose we have translational motion of a fronto-parallel plane. The intensity of point on that plane (x, y) at time t is expressed as $I(x, y, t)$, and this plane is undergoing a translation motion (u, v), such that at all locations x, y:

$$I(x, y, t) = I(x + u, y + v, t + 1). \tag{2}$$

When motion is assumed to be constant over the entire image and the brightness constancy assumption applies [11], the motion parameters (u, v) can be estimated by solving a linear system. The constraints for each pixel (x, y) are of the form:

$$-I_t(x, y) = I_x(x, y)u + I_y(x, y)v, \tag{3}$$

where I_x, I_y, I_t are the spatio-temporal image derivatives. Now, suppose our measurements of the function I are not samples of the value at or near a pixel x, y, but rather a more general spatial sampling of the function. Using the sampling described in (1), and differentiating $\hat{I}(p_i)$ with respect to time we get:

$$\begin{aligned} \hat{I}_t(p_i) &= \frac{\delta}{\delta t} \iint I(r) w_i(r) dr \qquad (4) \\ &= \iint I_t(r) w_i(r) dr \\ &= \iint -(I_x(r)u + I_y(r)v) w_i(r) dr \\ &= -u \iint I_x(r) w_i(r) dr - v \iint I_y(r) w_i(r) dr \end{aligned}$$

This seems promising, an estimate of the temporal derivative $\hat{I}_t(p_i)$ is simple to obtain but, since we are no longer densely sampling I, we cannot easily estimate I_x or I_y. However, given that w_i is compactly supported, this is equivalent to:

$$\begin{aligned} &\hat{I}_t(p_i) = \\ &\quad -u \textstyle\iint I(r) w_{ix}(r) dr - v \iint I(r) w_{iy}(r) dr \end{aligned}$$

and careful camera design can make it feasible to estimate these integral terms by, for instance, comparing the response of two integral pixels whose weight functions differ only in a slight shift in the x or y direction. Defining $w_{i\delta x}, w_{i\delta y}$ as such filters, we get:

$$\begin{aligned} &\hat{I}_t(p_i) = \\ &\quad -u \left(\textstyle\iint I(r) w_i(r) dr - \iint I(r) w_{i\delta x}(r) dr \right) \\ &\quad -v \left(\textstyle\iint I(r) w_i(r) dr - \iint I(r) w_{i\delta y}(r) dr \right), \end{aligned}$$

which is now a constraint on u, v based only on integral measurements. Estimating motion with these constraints is straightforward and computationally inexpensive; experimental results are given in Section 4.1. In the next section, we show how to incorporate known image motion, obtained by external means, into a compressed-sensing reconstruction problem using these constraints.

3. COMPRESSED-SENSING RECONSTRUCTION WITH KNOWN MOTION

We consider the problem of reconstructing the video of a moving scene given integral-pixel measurements from a camera with known motion. Our algorithm is based on the compressed-sensing image-reconstruction framework. While we describe the algorithm using translational motion, it can be generalized to any motion for which partial derivatives with respect to the motion parameters can be estimated.

It is well known that a vectorized natural image I can *usually* be factored as $I = Bc$ where B is a basis matrix and c is a sparse vector. Although many choices are possible, it is common for the basis matrix B to represent a complete wavelet basis [12] such that each column of B is a single wavelet.

We describe a set of integral measurements as a matrix multiplication,

$$\hat{I} = M^{\mathsf{T}} B c, \tag{5}$$

where columns of the measurement matrix M correspond to weight functions in an integral-pixel camera.

Compressed-sensing theory states that, subject to technical conditions on $M^{\mathsf{T}} B$, it is possible to accurately estimate the original image with a measurement set of cardinality proportional to the number of non-zero elements in the sparse image representation c. The important step is to solve for the sparsest c that satisfies the constraints imposed by the image measurements. Many methods have been proposed to estimate c, in this work we use a method based on conjugate-gradient descent [13]. Once the sparse representation c is known, reconstructing the image is a simple matrix multiplication.

In extending the model to video the following parametrization has been proposed [10]:

$$\begin{aligned} \hat{I}_1 &= M_1^{\mathsf{T}} B c_1 \\ \hat{I}_2 &= M_2^{\mathsf{T}} B c_2 \end{aligned}$$

Here we consider only a two-frame 'video' but the model naturally generalizes to longer sequences of images. Two meth-

[1] Note that the weight function may be required to be non-negative for a physical implementation. In this case, it may be possible to measure the positive and negative components of w_i separately and combine these values after sampling.

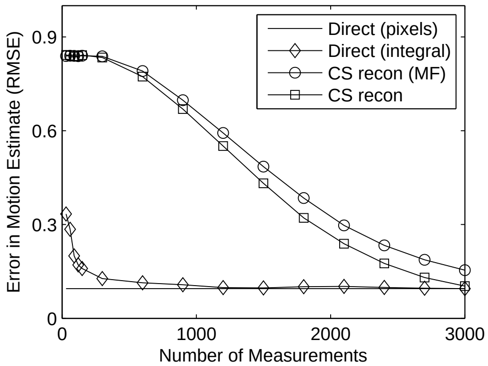

Fig. 1: Motion estimation error for several methods. The method we propose, *direct (integral)*, provides more accurate estimates than the reconstruction-based methods and requires significantly fewer measurements than the traditional full image method, *direct (pixels)*.

ods to perform image reconstruction for this model are proposed in the previous work. The first is to ignore the temporal constraints and solve separately for each image. The second method [10] uses the temporal constraints by solving simultaneously for both images. This method does not use the known motion parameters and instead depends on the temporal image derivative being sparse with respect to the sparsifying basis B.

Our method uses a differential approximation to constrain the motion between the image pair. This constraint enables us to reparametrize the problem in terms of a single sparse vector $\hat{c}$:

$$\hat{I}_1 = M_1^{\mathsf{T}}(B\hat{c} + \frac{\delta_x}{2}\frac{\partial B\hat{c}}{\partial x} + \frac{\delta_y}{2}\frac{\partial B\hat{c}}{\partial y}) \tag{6}$$

$$\hat{I}_2 = M_2^{\mathsf{T}}(B\hat{c} - \frac{\delta_x}{2}\frac{\partial B\hat{c}}{\partial x} - \frac{\delta_y}{2}\frac{\partial B\hat{c}}{\partial y}) \tag{7}$$

The image $B\hat{c}$ can be considered the image that would have been between images I_1 and I_2. Given known motion parameters δ_x, δ_y solving for a sparse $\hat{c}$ is equivalent to a standard compressed-sensing problem. The following derivation shows how to convert (6) into the same form as (5):

$$\hat{I}_1 = M_1^{\mathsf{T}}(B + \frac{\delta_x}{2}\frac{\partial B}{\partial x} + \frac{\delta_y}{2}\frac{\partial B}{\partial y})\hat{c}$$

$$\hat{I}_1 = M_1^{\mathsf{T}}\hat{B}\hat{c}$$

The advantage of our method is that the reconstruction algorithm uses the same number of measurements (constraints on the solution) but has fewer non-zero terms to estimate (the sparse vector $\hat{c}$ has roughly half the number of non-zero elements as in the pair of sparse vectors c_1, c_2 in the original parametrization). Experimental results in Section 4.2 show that this method gives more accurate image reconstructions with fewer measurements than the previous work.

4. EXPERIMENTS

4.1. Comparing Direct and Reconstruction-Based Motion Estimation

We evaluated four image motion estimation methods: Lucas-Kanade [14] translation estimation on the original images, our integral-pixel method, and two methods based on compressed sensing. The compressed-sensing methods first reconstructed the images (see the previous section for further details) and then estimated motion using the Lucas-Kanade method.

To construct the evaluation image set, single images were chosen randomly from the Middlebury optical flow evaluation dataset [15] and resized to 64×64 pixels. From each image a second image was generated by applying a small translation (u, v) (selected UAR such that $\|(u, v)\|_\infty < 1$ pixel).

Figure 1 shows the average error in the motion estimate as the number of image measurements changes (this number is fixed at $2 \times 64 \times 64$ for the *pixels* method). The compressed-sensing reconstruction methods used unconstrained IID Gaussian random matrices. The integral-pixel method used constrained matrices with one third of the measurements being random Gaussian and the remainder corresponding to x- and y- derivative measurements. For each experiment the same measurement matrix was used for both images.

The results show that estimating image motion directly from integral measurements (*direct (integral)*) gives substantially more accurate estimates than the reconstruction-based methods using the same number of measurements. The direct methods are also significantly more computationally efficient (in our experiments on average 500 times faster). This highlights that compressed-sensing reconstruction methods are not ideally suited for motion estimation.

4.2. Compressed-Sensing Reconstruction with Known Motion

In this section, we compare three image-reconstruction methods: our proposed method, a standard compressed-sensing method, and a multi-frame compressed-sensing method [10] We used the image set from Section 4.1 for evaluation. Image measurements were random Gaussian noise matrices and each image was sampled with a unique matrix. The x-axis of the plot shows the total number of measurements. For a sparsifying basis B we used the Daubechies wavelet basis [12]. Optimization for all methods was performed using GPSR [13]. All methods required roughly the same time to complete.

Figure 2 shows that reconstruction using known motion gives significantly lower reconstruction error. For example, we can achieve the same error level with either 600 measurements with known motion or approximately 1200 measurements with other methods. Perhaps the most surprising result is the lack of improvement from the multi-frame compressed-sensing reconstruction [10]. We conjecture that the multi-

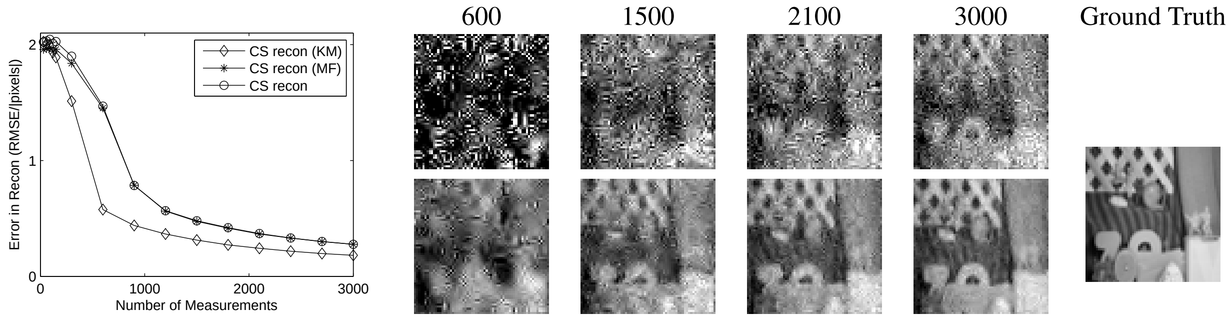

Fig. 2: Image-reconstruction error using known motion (KM) gives lower error reconstructions with significantly fewer measurements than traditional single-frame and multi-frame (MF) reconstructions.

frame method is unable to account for image shifts. Previous results using the method were only demonstrated using a static scene with small moving objects.

5. CONCLUSION

Explicit reasoning about image motion is important, but largely overlooked in the compressive-sensing literature. We have demonstrated that for certain motion estimation tasks, compressive measurements can provide more accurate motion estimation than a compressed-sensing reconstruction followed by traditional pixel-based motion estimation. We have also shown that inclusion of known motion into a compressed-sensing reconstruction can improve the quality of video reconstruction. These results demonstrate the importance of explicitly considering image motion in the context of compressive sensing. This also points towards the need for future work considering richer motion models and real-world implementations.

6. REFERENCES

[1] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, Ting Sun, K.F. Kelly, and R.G. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, March 2008.

[2] R. Fergus, A. Torralba, and W. T. Freeman, "Random lens imaging," Tech. Rep. MIT CSAIL TR2006-058, Massachusetts Institute of Technology, 2006.

[3] Ramesh Raskar, Amit Agrawal, and Jack Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," in *Proc. ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2006.

[4] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, 2008.

[5] Jon P Curzan, Christopher R Baxter, and Mark A Massie, "Variable-acuity imager with dynamically steerable programmable superpixels," in *Proceedings of SPIE, Infrared Technology and Applications*, 2003, vol. 4820.

[6] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, 2008.

[7] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.*, vol. 26, no. 3, 2007.

[8] Anat Levin, Rob Fergus, Frédo Durand, and William T Freeman, "Image and depth from a conventional camera with a coded aperture," in *Proc. ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2007.

[9] Volkan Cevher, Aswin C. Sankaranarayanan, Marco F. Duarte, Dikpal Reddy, Richard G. Baraniuk, and Rama Chellappa, "Compressive sensing for background subtraction," in *European Conference on Computer Vision*, 2008.

[10] Roummel F. Marcia and Rebecca M. Willett, "Compressive coded aperture video reconstruction," in *European Signal Processing Conference (EUSIPCO)*, 2008.

[11] B. K. P. Horn, *Robot Vision*, McGraw Hill, New York, 1986.

[12] Ingrid Daubechies, *Ten Lectures on Wavelets*, Soc for Industrial & Applied Math, Dec. 1992.

[13] M.A.T. Figueiredo, R.D. Nowak, and S.J. 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, Dec. 2007.

[14] Simon Baker and Iain Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, 2004.

[15] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," in *IEEE International Conference on Computer Vision*, Oct. 2007.