Report Number: wucse-2009-16

2009

On Unusual Pixel Shapes and Image Motion

Nathan Jacobs, Stephen Schuh, and Robert Pless

We introduce the integral-pixel camera model, where measurements integrate over large and potentially overlapping parts of the visual field. This models a wide variety of novel camera designs, including omnidirectional cameras, compressive sensing cameras, and novel programmable-pixel imaging chips. We explore the relationship of integral-pixel measurements with image motion and find (a) that direct motion estimation using integral-pixels is possible and in some cases quite good, (b) standard compressive-sensing reconstructions are not good for estimating motion, and (c) when we design image reconstruction algorithms that explicitly reason about image motion, they outperform standard compressive-sensing video reconstruction. We show experimental... **Read complete abstract on page 2.**

On Unusual Pixel Shapes and Image Motion

Nathan Jacobs, Stephen Schuh, and Robert Pless

Complete Abstract:

We introduce the integral-pixel camera model, where measurements integrate over large and potentially overlapping parts of the visual field. This models a wide variety of novel camera designs, including omnidirectional cameras, compressive sensing cameras, and novel programmable-pixel imaging chips. We explore the relationship of integral-pixel measurements with image motion and find (a) that direct motion estimation using integral-pixels is possible and in some cases quite good, (b) standard compressive-sensing reconstructions are not good for estimating motion, and (c) when we design image reconstruction algorithms that explicitly reason about image motion, they outperform standard compressive-sensing video reconstruction. We show experimental results for a variety of simulated cases, and have preliminary results showing a prototype camera with integral-pixels whose design makes direct motion estimation possible.

2009-16

On Unusual Pixel Shapes and Image Motion

Authors: Nathan Jacobs, Stephen Schuh, Robert Pless

Corresponding Author: jacobsn@gmail.com

Abstract: We introduce the integral-pixel camera model, where measurements integrate over large and potentially overlapping parts of the visual field. This models a wide variety of novel camera designs, including omnidirectional cameras, compressive sensing cameras, and novel programmable-pixel imaging chips. We explore the relationship of integral-pixel measurements with image motion and find (a) that direct motion estimation using integral-pixels is possible and in some cases quite good, (b) standard compressive-sensing reconstructions are not good for estimating motion, and (c) when we design image reconstruction algorithms that explicitly reason about image motion, they outperform standard compressive-sensing video reconstruction. We show experimental results for a variety of simulated cases, and have preliminary results showing a prototype camera with integral-pixels whose design makes direct motion estimation possible.

Type of Report: Other

On Unusual Pixel Shapes and Image Motion

Nathan Jacobs, Stephen Schuh, Robert Pless
Washington University
St. Louis, MO, 63130 USA
(jacobsn|schuhs|pless)@cse.wustl.edu

Abstract

We introduce the integral-pixel camera model, where measurements integrate over large and potentially overlapping parts of the visual field. This models a wide variety of novel camera designs, including omnidirectional cameras, compressive sensing cameras, and novel programmable-pixel imaging chips. We explore the relationship of integral-pixel measurements with image motion and find (a) that direct motion estimation using integral-pixels is possible and in some cases quite good, (b) standard compressive-sensing reconstructions are not good for estimating motion, and (c) when we design image reconstruction algorithms that explicitly reason about image motion, they outperform standard compressive-sensing video reconstruction. We show experimental results for a variety of simulated cases, and have preliminary results showing a prototype camera with integral-pixels whose design makes direct motion estimation possible.

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 [8], or to approximate large parts of the 4D light-field incident on a camera [9]. 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. Motion is an especially important concept for integral-pixels, because they capture light over a large region and therefore offer the potential for constructing cameras that operate at higher frame rates and/or with lower pixel noise.

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 arbitrary integral-pixel cameras. Second, we designed and built a novel camera where each pixel integrates over variously oriented, long narrow receptive fields, and show that this camera is effective at estimating translations with relatively few integral measurements compared to standard pixel.

Third, 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. Finally, we discuss why it may be difficult to directly approach more general structure-from-motion problems within the integral-pixel camera model.

1.1. Related Work

Within the vision community, recent work in novel camera design has mostly focused on omnidirectional and panoramic cameras. Describing the imaging geometry of such cameras is usually done by first defining whether all the rays imaged by the camera pass through a single point (a central projection camera) or not. For central projection cameras, the calibration specifies the mapping between image pixels and rays in space. For non-central projection cameras, this mapping must include both the direction of the ray and some point along that ray, often specified in terms of caustics.

Although this was mostly ignored in the subsequent work, the initial raxel model of generalized camera geometry [11] models both the direction of the ray, and the point spread function of the camera. Even point source of light may illuminate a small region of pixels—conversely, light measured at a pixel may actually come from a collection of rays in the neighborhood if the dominant raxel direction. Recently the cone-pixel model [7] was presented to model omnidirectional cameras with spatially-varying sampling. They present methods for calibration and motion estimation based on the geometric constraints defined by the intersection of cones from two views.

In contrast to work on ray-based camera models, much recent camera design work considers alternatives to sampling the scene by focusing on measurements that do not correspond to a simple ray or cone of the visual space. This includes capturing an image using time-coded shutters [15], coded combinations of location and wavelength [17], and compressive sensing [3, 8, 5]. Other work includes the use of coded apertures [16, 13].

These approaches seek to reconstruct approximations of the (image, hyper-spectral data cube, or light-field) as if they were sampled by a regular pixel grid. Although two papers very recently consider the problem of independent motion detection [4], and reconstructing multiple frames of a video [14], both focus on reconstructing the difference images or the video frames rather than estimating the frame-to-frame motion.

A number of unique camera designs have been created to sample the 4D light-field. The random lens imaging camera [9], places an array of randomly oriented mirrors in the optical path so that each pixel samples a "random" (but known) ray of the local 4D light-field. One of the earliest works [1] to directly capture the 4D light-field used a similar camera design to ours with an additional lens in front of the lenticular array. In their work they use the lenticular array as a set of lenses so that each pixel samples a different part of the 4D light-field. Our work uses the lenticular array as a diffuser to integrate across a set of viewing directions.

2. The Integral-Pixel Imaging Model

We define the integral-pixel imaging model in relation to a standard ray-based camera model. Let $I(r)$ be the response of a standard camera at particular ray/pixel r. Within the more general integral-pixel imaging model, we consider a camera whose pixels capture an integral of an arbitrary, non-negatively weighted, region of the imaging sphere. 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, with values ranging from zero to one, that describes how pixel i samples the view sphere. For the traditional compact pixel model $w_i(r)$ is non-zero in a small region (generally less than a few pixel widths). For an integral pixel camera weight functions need only have compact support (*i.e.* vanish at infinity).

The remainder of this paper explores the relationship between motion estimation and integral-pixel cameras. The next section describes a method for directly estimating motion parameters using measurements from an integral-pixel camera.

3. Direct Motion Estimation with Integral-Pixels

We describe how to extend the Lucas-Kanade motion estimation method to work for integral-pixel cameras. This section addresses the case of translational motion; Section 7 discusses issues with extending to richer motion models.

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:

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

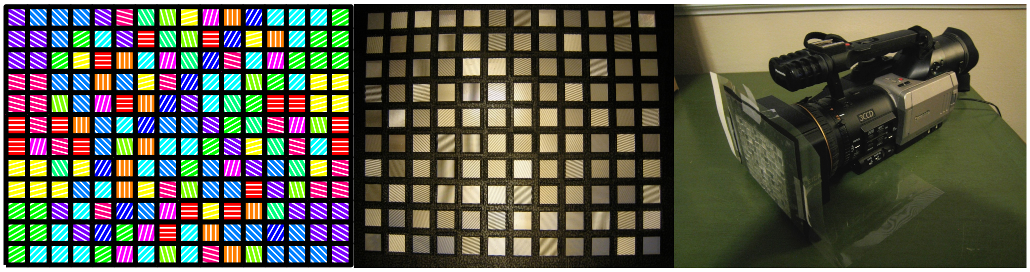
When motion is assumed to be constant over the entire image and the brightness constancy assumption applies [12], the motion parameters (u, v) can be estimated by solving a linear system. Constraints in the system are generated for each pixel (x, y) as follows:

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

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



(a) (b) (c)

Figure 1: We constructed an integral-pixel camera using a grid of oriented lenticular arrays. (a) The template for an optimal tiling pattern. (b) The physical instantiation of the optimal tiling pattern. (c) The prototype video camera using the lenticular tiling pattern.

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:

$$\hat{I}_t(p_i) = u \iint I(r) w_{ix}(r) dr + v \iint I(r) w_{iy}(r) dr.$$

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:

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

which is now a constraint on u, v based only on integral measurements. In the next section, which shows the design and calibration of such a camera, we recognize that because the weight functions are for a real system pixels and are empirically determined, there may not be an identical but slight shifted weight function, and for each w_i we solve for approximate $w_{i\delta x}, w_{i\delta y}$ as a weighted combination of nearby measurements.

4. A Prototype Camera Design

We constructed an integral-pixel video camera (see Figure 1) using a standard video camera and a grid of lenticular lenses. The lenticular lenses we use consist of an array of cylindrical magnifying lenses. They are commonly used to create the illusion of motion or depth on a plane as the viewpoint changes. We use lenticular lenses because when viewed in the proper direction they sample interesting integrals[1] of the viewing sphere.

[1] Cylindrical lenticular lenses measure approximate line integrals perpendicular to the orientation of the cylindrical lenses.

The design consists of lenticular tiles, approximately 1cm^2, at 12 orientations. These tiles are arranged in a regular grid between two plastic adhesives sheets, as shown in Figure 1b. This assembly is then attached to the front of a standard video camera. The camera is focused so that the lenticular array lies exactly at the focal plane, so each pixel of the camera captures light that passes through a small part of the lenticular array. Images captured from this camera are shown in Figure 4.

The tiling pattern was chosen, using an integer linear solver, so that no orientation was undersampled in the perpendicular direction to the cylindrical lenses. This can be seen in Figure 1a where, for example, there is red (horizontal) array in each column and an orange (vertical) array in each row. Each tile is made from a 75 lines-per-inch lenticular sheet—which is small enough that the fraction of the tile imaged by each pixel covers more than period of the lenticular repetitions. The ray diagram in Figure 2 shows how one pixel samples the world.

Each of the 720×480 pixels in the video camera now gives a different integral and, because of the relatively large size of the tiles compared to the integration area of a single video camera pixel, neighboring pixels have shifted integration patterns. This enables computation of estimates of the partial derivative of the image measurements needed for motion estimation using the direct motion estimation method described in Section 3.

4.1. Calibration

The goal of the calibration step is to solve for the weight function $w_i(x)$ and the partial derivatives of $w_i(x)$ for each pixel in the video camera. By calibrating each pixel we avoid the issues of barrel distortion in the camera and we capture any non-uniformity.

To solve for the weight functions $w_i(x)$ we placed the camera so that it was imaging a computer LCD monitor. We then displayed images on the monitor and captured the

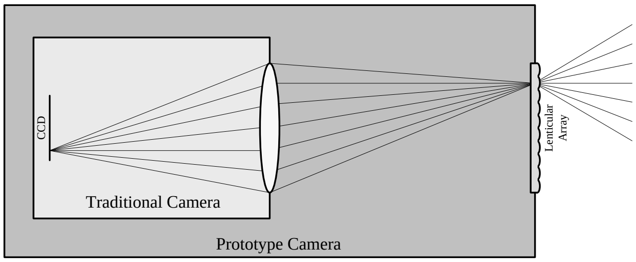


Figure 2: A ray diagram of the rays in space (on the far right) that affect the measurement at a single pixel. This spread is limited to one plane, in this case, the plane of the page due to the orientation of the lenticular tile.

corresponding images from the camera. We used a set of 900 simple calibration patterns; an example of the calibration pattern and the corresponding image from the prototype camera can be seen in Figure 4. For a given pixel solving for the weight function $w_i(x)$ is trivial because of the simplicity of the calibration pattern. Least-squares would suffice for more general calibration patterns. Although we did not experience this problem, more general calibration patterns (with more white pixels) could be helpful if low-light levels are a problem.

Next we must find a way to estimate the partial derivatives of the measurements obtained from each weight functions $w_i(x)$. We can compute the true partial derivatives by computing the gradients of $w_i(x)$. Ideally we would measure the image with this but we cannot because it does not exist and has negative values; we have to estimate the desired measurement. To obtain an estimate of this partial we solve for the linear combination of the measurements at pixels near pixel x (we used a 10×10 pixel neighborhood) that best approximate the true partial. Using this method is more robust than simply subtracting the value of the neighboring (as proposed in Section 3) because it can account for the variability due to construction and image distortion.

Figure 4 contains samples of the weight function for three different video camera pixels. Examples of motion estimation for this camera and using this calibration are given in Section 6.2. The next section shifts from the problem of motion estimation to the problem of reconstructing a video with measurements from an integral-pixel camera.

5. 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 [6] such that each column of B is a single wavelet.

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

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

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 [10]. 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 [14]:

$$\hat{I}_1 = M_1^{\mathsf{T}} Bc_1$$
$$\hat{I}_2 = M_2^{\mathsf{T}} Bc_2$$

Here we consider only a two-frame 'video' but the model naturally generalizes to longer sequences of images. Two methods 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 [14] 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 stan-

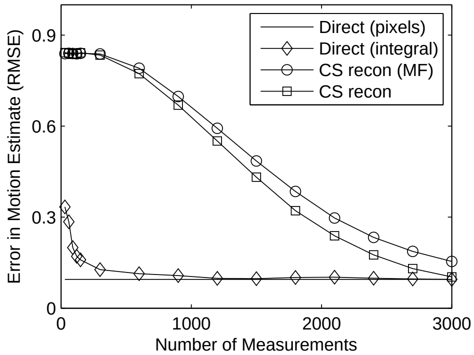


Figure 3: Motion estimation error for several methods. The method we propose, *direct (integral)*, provides more accurate estimates than the reconstruction-based methods and requires significantly fewer measurements than the traditional full image method, *direct (pixels)*.

dard 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 6.3 show that this method gives more accurate image reconstructions with fewer measurements than the previous work.

6. Experiments

6.1. Comparing Direct and Reconstruction-based Motion Estimation

We evaluated four image motion estimation methods: Lucas-Kanade translation estimation on the original images, our integral-pixel method, and two compressed sensing based methods. 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 [2] and resized to 64×64 pixels. From each image a second image was generated by applying a small translation t (selected UAR such that $\|t\|_\infty < 1$ pixel).

The plot in Figure 3 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 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 style measurements and the associated reconstruction methods are not ideally suited for motion estimation.

6.2. Motion Estimation for the Prototype Camera

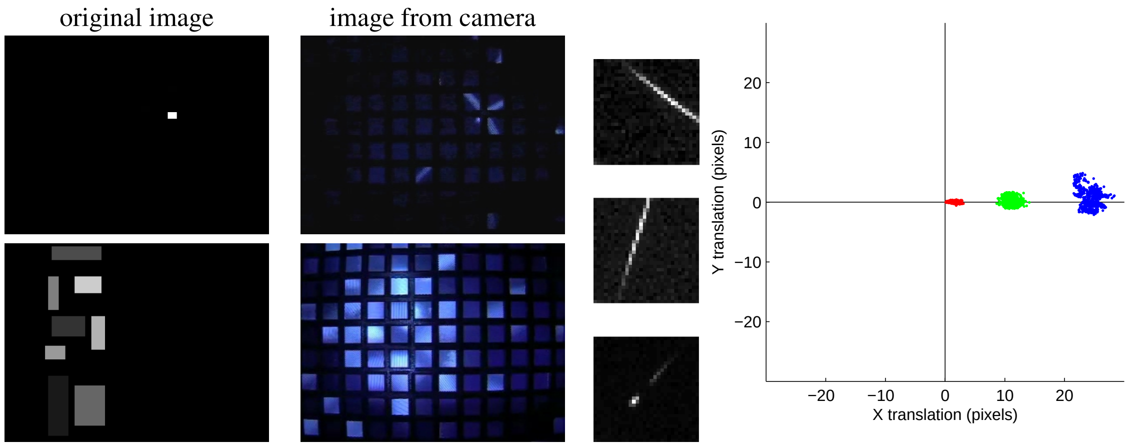
We tested translational motion estimation for the prototype camera described in Section 4. We used three videos with the same pattern moving at a constant velocity to the right. Figure 4 shows an example frame from the test videos and a frame from the prototype camera (the integral measurements). To obtain measurements we used the same setup as the calibration case.

Figure 4 shows motion estimation results from our prototype camera. Each point represents a motion estimate for a given frame. The method gives consistent motion estimates for the three different motions. These results demonstrate the possibility of obtaining an estimate of motion using our prototype camera. We attribute the uncertainty in the motion estimates to the limited resolution of our calibration routine. Increasing this resolution would significantly increase the accuracy of these estimates.

6.3. Compressed Sensing Reconstruction with Known Motion

We evaluated image reconstruction methods using the image set used in the known motion experiment in Section 6.1. The methods compared are our proposed method, a standard compressed sensing reconstruction method, and a multi-frame compressed sensing method [14].

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 [6]. Optimization for all methods was performed using GPSR [10], a fast gradient-projection based method. All methods required roughly the same time to complete.



(a) Example frames

(b) Example weight functions

(c) Motion estimates

Figure 4: (a) Frames from the calibration (top) and translational motion test video (bottom). The original images are at the left and the corresponding images captured by our prototype camera (shown in Figure 1) are at right. (b) Weight functions $w_i(r)$ solved for during calibration for three different pixels from our prototype camera. (c) Motion estimation results obtained using our prototype camera. Each point corresponds to the estimate for a frame of a video with constant motion to the right (different colors/colors correspond to different temporal subsamplings of the video).

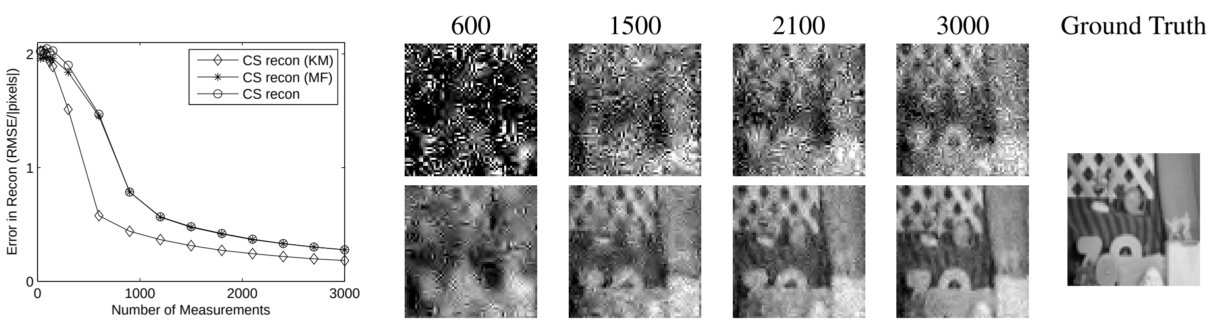


Figure 5: Image reconstruction error using known motion (KM) gives lower error reconstructions with significantly fewer measurements than traditional single-frame and multi-frame (MF) reconstructions.

The results in Figure 5 show 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 [14]. We conjecture that the multi-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.

7. Discussion—Challenges in Approaching Full Structure-from-Motion

Given that we have found structured integral-pixel measurements which are better at estimating motion than typical pixels or typical compressive sensing approaches, it is natural to ask what sampling strategy may be effective for computing more complicated image motions, such as affine, general-linear, or projections of arbitrary rigid 3D motions.

Some of the challenges are visible directly in the affine case. Assuming affine camera motion and starting from (4)

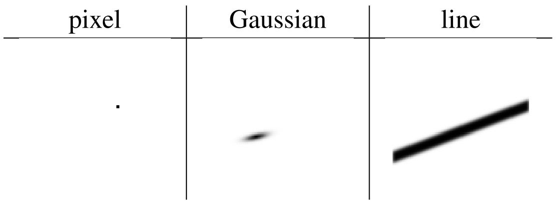


Figure 6: Examples of the filters used to generate measurements for Figure 7.

we have:

$$\hat{I}_t(p_i) = -\iint (I_x(r)(ax+by+c)+I_y(r)(dx+ey+f))w_i(r)\,dr \tag{8}$$

The issue is that there is no way, in general, to measure or estimate any of the four quantities with form similar to:

$$\iint x(r)I_x(r)w_i(r)\,dr \tag{9}$$

The difficulty arises because of the $x(r)$ and $y(r)$ inside the integral. This issue is not present in when using compact pixels because it is reasonable to assume that $x(r)$ and $y(r)$ are known and constant over the region where w_i is non-zero. If $x(r)$ and $y(r)$ are constant they can be moved outside the integral leaving an integral that is easy to measure or approximate.

7.1. The Ray Approximation

For integral-pixels and affine motion a key difficulty is in estimating (9). A solution to this issue is to use the center-of-mass of the filter as an estimate, more specifically $x_i = \iint x(r)w_i(r)\,dr / \iint w_i(r)\,dr$ and similarly for y_i. This allows equations for the form $\iint x(r)I_x(r)w_i(r)\,dr$ to be rewritten as $x_i \iint I_x(r)w_i(r)\,dr$. Using this approximation we can now construct the linear system of equations needed to solve for the motion parameters. This reduces the camera model to a ray-based camera model. The next section explores the implications of this approximation on the accuracy of motion parameter estimates.

7.2. Evaluating the Ray Approximation for Motion Estimation

To evaluate the impact of the ray approximation we conducted a simulation study. For the study, we constructed two sets images from a single base image from the Middlebury optical flow evaluation dataset [2]. The first set consists of the base image and a set of the base images subject to known small affine deformation. Motions were selected uniformly at random such that the deformation was less than two pixels at the corners of the 160×120 image. The second set of image pairs was constructed by modifying the contrast of the first set. Specifically, the maximum brightness range was decreased quadratically from 256 at the bottom of the image to zero at the top.

We compare three different weight functions: a Dirac weight function (pixels), a line integral, and an oriented Gaussian. Each trial used 50 integral measurements of a given type at different locations and orientations. We assume that the weight functions, $w_i(x)$, and the partial derivatives of these measurements with respect to the motion parameters are known. We used these measurements and the ray approximation to estimate motion using our direct motion estimation method (see Section 3).

Figure 7 shows errors in estimating motion parameters for the three weight functions. In each plot, the points correspond to the difference between the ground truth motion and the estimated motion for a single affine motion parameter. The ellipses represent the one standard deviation isocontour of the errors. The results highlight several aspects of the performance characteristics of motion estimation using integral-pixels. First, despite the ray approximation the oriented Gaussian blobs outperform single pixel measurements in all cases. Second, for the case of scale motion and the *lines* measurements the error for the y component of the motion has greater error than the x component. This is due to the nonuniformity of contrast in the y direction. Third, the *lines* measurements are very accurate for the translation case. The *lines* measurements are accurate, in part, because for the translation case the ray approximation is unnecessary.

These results show that the ray approximation is not suitable for general motion estimation for integral-pixel cameras. Hence, for an interesting and important class of new physical camera designs existing ray-based camera models do not accurately describe the relationship between motion and image measurements.

8. Conclusion

We have presented a novel imaging model, the integral-pixel model, that accurately describes a new and important class of cameras. We have demonstrated through simulations that for simple motion estimation tasks, integral-pixels can provide better motion estimation that the same number of standard pixels. We developed a prototype camera where each pixel optically integrates light from large regions of the view sphere, and have shown that we can calibrate this camera sufficiently to estimate simple motions. We have also shown that inclusion of known motion into the compressed sensing framework can improve the quality of video reconstruction.

This paper explored the relationship between scene motion and the measurements made by integral-pixel cameras. We demonstrated theoretically and empirically that ray-based models cannot accurately describe this relationship for integral-pixel cameras. We believe the integral-

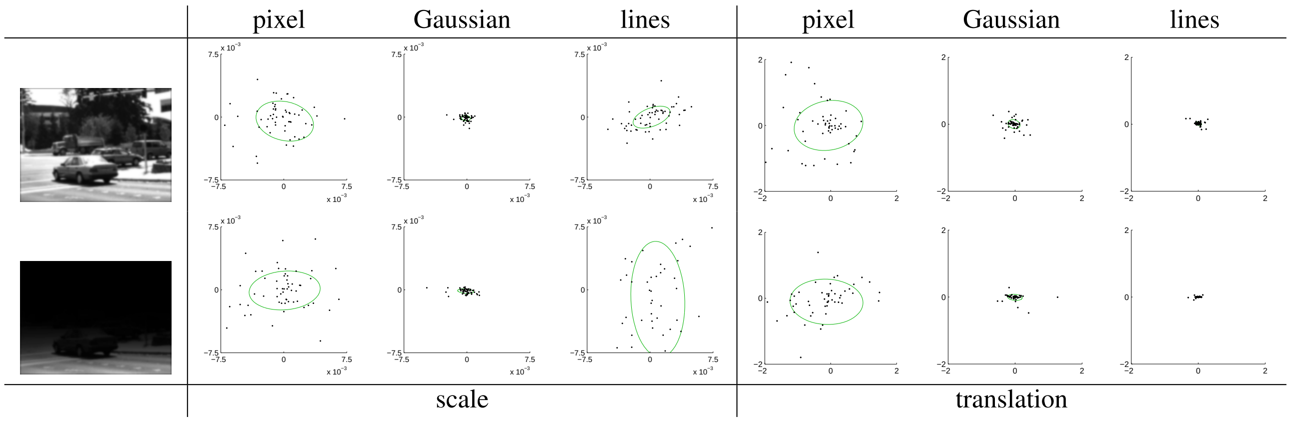


Figure 7: Estimation error for simulations using three types of weight functions and two different images. See Figure 6 for examples of the corresponding filters. The first, unsurprising, observation is that using traditional single-pixel measurements is less accurate than using elongated Gaussian measurements. Apparently the elongation was not sufficient to cause the ray approximation to bias the motion estimate. Using *lines* measurements results in very accurate translational motion estimation but inaccurate scale estimation (especially in the non-uniform contrast case).

pixel model is important and furthering our understanding the relationship between this model and motion is an important area of future work.

References

[1] E. Adelson and J. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):99–106, 1992. 2

[2] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proc. IEEE International Conference on Computer Vision*, pages 1–8, Oct. 2007. 5, 7

[3] 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*, 25(2):21–30, 2008. 2

[4] V. Cevher, A. C. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa. Compressive sensing for background subtraction. In *Proc. European Conference on Computer Vision*, pages 155–168, 2008. 2

[5] J. P. Curzan, C. R. Baxter, and M. A. Massie. Variable-acuity imager with dynamically steerable programmable superpixels. In *Proceedings of SPIE, Infrared Technology and Applications XXVIII*, volume 4820, page 318, 2003. 2

[6] I. Daubechies. *Ten Lectures on Wavelets (C B M S - N S F Regional Conference Series in Applied Mathematics)*. Soc for Industrial & Applied Math, December 1992. 4, 5

[7] T. Debaecker, R. Benosman, and S.-H. Ieng. Cone of view camera model using conformal geometric algebra for classic and panoramic image sensors. In *IEEE Workshop on Omnidirectional Video*, 2008. 2

[8] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):83–91, March 2008. 1, 2

[9] R. Fergus, A. Torralba, and W. T. Freeman. Random lens imaging. Technical Report MIT CSAIL TR 2006-058, Massachusetts Institute of Technology, 2006. 1, 2

[10] 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*, 1(4):586–597, Dec. 2007. 4, 5

[11] M. D. Grossberg and S. K. Nayar. The raxel imaging model and ray-based calibration. *International Journal of Computer Vision*, 61(2):119–137, 2005. 2

[12] B. K. P. Horn. *Robot Vision*. McGraw Hill, New York, 1986. 2

[13] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *Proc. ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, page 70, 2007. 2

[14] R. Marcia and R. Willett. Compressive coded aperture superresolution image reconstruction. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 833–836, 2008. 2, 4, 5, 6

[15] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *Proc. ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 795–804, New York, NY, USA, 2006. ACM. 2

[16] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graph.*, 26(3):69, 2007. 2

[17] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47(10):B44–B51, 2008. 2