# NON-PARAMETRIC ACQUISITION OF NEAR-DIRAC PIXEL CORRESPONDENCES

Bradley Atcheson and Wolfgang Heidrich
*Department of Computer Science, The University of British Columbia, Vancouver, Canada*

Keywords: Pixel Correspondences, Bloom Filter, Environment Matting, Structured Light.

Abstract: Many computer vision and graphics applications require the acquisition of correspondences between the pixels of a 2D illumination pattern and those of captured 2D photographs. Trivial cases with only one-to-one correspondences require only a few measurements. In more general scenes containing complex inter-reflections, capturing the full reflectance field requires more extensive sampling and complex processing schemes. We present a method that addresses the middle-ground: scenes where each pixel maps to a small, compact set of pixels that cannot easily be modeled parametrically. The coding method is based on optically-constructed Bloom filters and frequency coding. It is non-adaptive, allowing fast acquisition, robust to measurement noise, and can be decoded with only moderate computational power. It requires fewer measurements and scales up to higher resolutions more efficiently than previous methods.

## 1 INTRODUCTION

Many problems in computer vision require the establishment of correspondences between camera pixels and either a single or multiple points on scene objects or illuminants. For example, in 3D scanning it is common to project a sequence of light stripes or encoded patterns onto an object in order to reconstruct the geometry via the observed displacement of projector pixels. In these settings, each camera pixel receives only contributions from a single point on the illuminant, i.e. the point spread function (PSF) is a Dirac peak. Binary encodings such a Gray codes (Bitner et al., 1976) can solve this pixel correspondence problem. In practice however, they suffer from errors since the PSFs are rarely perfectly Dirac, and such binary codes do not readily admit subpixel-accurate correspondences. At the other extreme, various simplifications of the full 8D reflectance field (Debevec et al., 2000) can be employed to obtain the low frequencies.

In our work, we focus on the intermediate problem of small, *near-Dirac* point spread functions which must be captured with high subpixel precision. For such applications it is not only necessary to estimate small but finite-sized PSFs, but we must do so robustly, and with high accuracy. Due to their high-frequency anisotropic nature, a non-parametric description of the PSFs is preferable to the axis-aligned box (Zongker et al., 1999) or oriented Gaussian models (Chuang et al., 2000) used in environment matting.

Bloom filters are extremely space-efficient data structures for probabilistic set membership testing (Bloom, 1970). We show how such structures can be optically constructed in the context of the pixel correspondence problem, and then inverted using heuristics and compressive sensing algorithms. To this we add a frequency-based environment matting scheme (Zhu and Yang, 2004), but modified to increase efficiency. It naturally handles one-to-many pixel correspondences in a non-parametric fashion. The result is a combined binary/frequency-based non-adaptive coding scheme that requires a comparatively small number of input images while being robust under noise. Processing time is on the order of minutes on a desktop machine, which is significantly faster than the general light transport acquisition methods based on compressed sensing.

## 2 RELATED WORK

**Structured Light Scanning** applications typically employ efficient encodings such as Gray codes (Bitner et al., 1976) that require only a small number of images. For scanning moving objects, other codes have been developed which allow track-

Atcheson B. and Heidrich W..
NON-PARAMETRIC ACQUISITION OF NEAR-DIRAC PIXEL CORRESPONDENCES.
DOI: 10.5220/0003825002470254
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 247-254
ISBN: 978-989-8565-04-4

ing over time (Hall-Holt and Rusinkiewicz, 2001; Rusinkiewicz et al., 2002). These stripe encodings are efficient for the purpose of structured light scanning, but have shortcomings. They can only determine one-to-one pixel mappings. While acceptable for many 3D scanning purposes, the inability to deal with mixtures of pixels can result in artifacts.

Scharstein and Szeliski (2003) projected both Gray-coded stripes and sine waves of different spatial frequencies. They note that binary codes can be difficult to measure in the presence of low scene albedo or low signal-to-noise ratio and overcame this by projecting both the binary code and its inverse. In general though, binary codes are very robust. Methods based on absolute amplitude measurements are highly dependent upon accurate radiometric calibration and consistent scene albedo.

**Environment Matting** estimates the occlusion-free light transport matrix between a 2D background and camera image. Zongker et al. (1999) used binary stripe patterns both horizontally and vertically to obtain correspondences in the form of rectangular axis-aligned regions on the background for each camera pixel. This method suffers from ambiguities in cases where two disjoint regions on the background map to a single camera pixel Chuang et al. (2000) proposed a number of improvements to this algorithm, including one that generalizes the axis-aligned boxes to oriented Gaussian regions of influence, and one that resolves the bimodal distribution ambiguity via additional (potentially redundant) diagonal sweeps.

For specular correspondences with small spatial support, it is possible to derive algorithms that require significantly fewer images by employing learning approaches (Wexler et al., 2002) or even single images by optical flow (Atcheson et al., 2008).

Peers and Dutré (2003) proposed the use of wavelets as illumination patterns for environment matting. Their initial algorithm was adaptive, i.e. it required processing the results of captured images to decide which patterns to project next, drastically increasing the acquisition time. This disadvantage was remedied in a later work (Peers and Dutré, 2005), in which the authors use sparsity priors to project results obtained with a fixed set of illumination patterns into a new wavelet representation. While this method produces excellent results for wide point spread functions, it is less applicable to sharp PSFs.

Zhu and Yang (2004) have proposed a temporal frequency-based coding scheme whereby the intensity of each pixel is set according to a 1D signal (a sinusoid). Our intra-tile coding scheme is based on this method but employs a second carrier, ninety degrees out of phase of the primary sinusoid, to double the information density at no extra cost. The use of only integral frequencies satisfies the Nyquist ISI criterion and allows for very fast, easy and robust DFT-based decoding. We choose to uniquely code individual pixels (within each tile) rather than coding whole rows and columns of the illuminant. This allows our method to scale up to higher illuminant resolutions, and to naturally handle PSFs of arbitrary (small) shape, rather than assuming a parametric form.

**Light Transport Matrix.** Recent papers have focused on the general problem of estimating the light transport matrix between illuminant and camera pixels. Most employ strategies similar to those used in environment matting. Sen et al. (2005) propose a hierarchical decomposition into non-interfering regions. The adaptive approach requires many images to resolve PSFs overlapping multiple regions.

Garg et al. (2006) note that the light transport matrix is often data-sparse. They exploit this, along with its symmetry due to Helmholtz reciprocity, in their adaptive acquisition algorithm that divides the matrix into blocks and approximates each with a rank-1 factorization. Wang et al. (2009) similarly seek a low-rank approximation to the full matrix. However, they do so by densely sampling rows and columns of the matrix (which requires a complex acquisition setup) and then using a kernel Nyström method to reconstruct the full matrix. These methods assume the matrix to be data-sparse (compressible).

Methods based on compressed sensing are beginning to appear. Sen and Darabi (2009) and Peers et al. (2009) both describe promising, non-adaptive, methods that transform the light transport into a wavelet domain in which it is more sparse. While these methods allow for capturing very complex light transport, they still require on the order of hundreds to thousands of images at typical resolutions, and many hours of decoding time to obtain results.

Our method combines advantages of many of the aforementioned works in that it is both scalable and robust, while being conceptually simple and easy to implement. For typical configurations we require on the order of a few hundred images that can be acquired non-adaptively in seconds and then processed in minutes on a standard desktop computer. Unlike more advanced light transport acquisition methods, we cannot acquire large, diffuse PSFs (one-to-many correspondences). But for the case of small, finite PSFs, those methods require many images to resolve high frequency detail. In contrast, our method efficiently captures accurate data at much lower cost in terms of acquisition and processing time.

# 3 ALGORITHM

We propose a combined binary/frequency-coded structured light pattern for estimating pixel correspondences. Appropriate acquisition setups are simple and inexpensive. All that is required is a spatially-addressable background illuminant (projector or LCD monitor), a camera and a reflective or refractive scene. Projected patterns are acquired by a synchronized camera and then decoded offline.

The detection algorithm is divided into two phases. First, the background is partitioned into small rectangular tiles (we use $8 \times 8$ pixels). Each tile is assigned a unique temporal binary code. A sequence of images is acquired where the tiles flash white or black according to their bit pattern. Since each camera pixel maps to a small area of the background, the measured signal consists of the superposition of these bit patterns. The task is then to determine which codes are present in the observed signal. We use sparsity and spatial coherence heuristics to solve it.

In the second phase we obtain per-pixel weights corresponding to the PSF. Each pixel within a tile is assigned a unique integral frequency and phase combination. We then acquire a sequence of patterns in which each pixel's intensity varies according to the amplitude of its corresponding sinusoidal waveform.

The first phase (inter-tile coding) may optionally use a frequency encoding similar to that of the second phase, but at higher resolution. We describe this method first in Section 3.1 and note that it performs very well in simulation. However, with real data that may not be subject to our simulated assumptions of additive white noise, we turn instead to the Bloom filter-based method described in Section 3.2. The second phase (intra-tile) is then described in Section 3.3.

## 3.1 Inter-tile Frequency Coding

As previously mentioned, we assign each tile a unique code. By enumerating tiles this way in 2D, we avoid the ambiguity suffered by methods that partition the background into rows and columns (Zongker et al., 1999; Chuang et al., 2000). In those schemes, a pixel containing contributions from rows $x_1 \neq x_2$ and columns $y_1 \neq y_2$ has four possible intersection points. The actual beam may have struck two, three or four of these points, and the natural way to eliminate the phantom points is to perform an additional scan pass using a different orientation (e.g. diagonal lines). However, for the unambiguous beams this pass is redundant and reduces efficiency.

The disadvantage of using 2D enumeration is that there are usually far more tiles requiring unique identifiers than either rows or columns. For example, a $1600 \times 1200$ monitor could be partitioned into 30,000 tiles of size $8 \times 8$. Were we to directly employ frequency-based environment matting (Zhu and Yang, 2004) on these, we would have a maximum frequency of 30kHz and thus require more than 60,000 captured images. Even the improvement we describe in Section 3.3 only halves this. But this does assume only integral frequencies and only two phases. We are in fact free to choose any appropriate frequency/phase sampling resolution. Figure 1 shows an example sampling lattice in frequency/phase parameter space. In the diagram a regular grid is used, with buffer regions in the very low and very high frequencies. Frequency estimation accuracy in these boundary regions is degraded, as predicted via the Cramér-Rao Bound (CRB), which places a lower bound on the variance of an unbiased estimator (Kay, 1993). While the CRB suggests that an optimal lattice would be nonuniformly spaced with frequency sampling density varying according to $f$, in practice the oscillations are small and we prefer a regular grid for simplicity. However, frequencies near 0 Hz and the Nyquist limit should nevertheless be avoided.

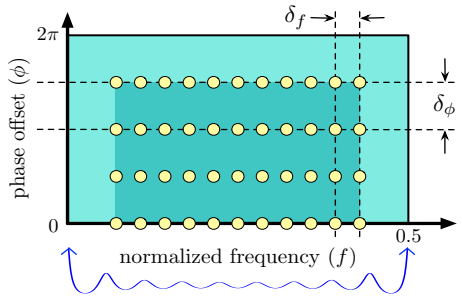



Figure 1: Sample points in frequency/phase space. $\delta_f$ and $\delta_\phi$ may be arbitrarily small. The graph below represents the Cramér-Rao Bound for the variance on frequency estimates. Note that accuracy degrades significantly near the 0 and 0.5 cycles/sample limits.

This very dense sampling requires a signal parameter estimation algorithm that can very accurately detect the frequencies. Periodograms, as used in the intra-tile coding step, are most useful when only integral frequencies are present. Otherwise, spectral leakage interferes. In higher resolution scenarios, better accuracy can be obtained via subspace methods such as ESPRIT (Roy and Kailath, 1989). Based on eigendecomposition of the signal covariance matrix, it is particularly well suited to the case of sinusoidal parameter estimation in a signal corrupted by additive white Gaussian noise.

Despite their great accuracy, subspace methods can fail when signals contain multiple components of very similar frequency. This is likely to occur if

we number the tiles in row- or column-wise order and map these directly to consecutive points in frequency/phase space, because many beams will strike near the tile boundaries and receive contributions from adjacent tiles. To ensure that spatial neighbors are not also frequency/phase-space neighbors it is necessary to label them according to a random, or low discrepancy sequence.

Our simulations in Section 4 indicate that 225,000 unique codes can be represented in 64 images. Unfortunately, real-world experiments could not reproduce these synthetic results. One possible explanation is that Gaussian noise is a poor model of the actual measurement noise and response-curve linearization error in our acquisition setup. While we believe that high resolution spectral methods show promise for pixel coding, our experiments suggest that too many images need be captured in order to obtain accurate estimates. For this reason we also developed the better-performing binary coding scheme described next.

### 3.2 Inter-tile Binary Coding

A set of N distinct tiles can easily by coded as consecutive natural numbers, whose binary representations require the acquisition of only $\log_2 N$ images. This scheme suffers from reliability problems, in that a single incorrectly-read bit can drastically alter the number. Gray codes ameliorate this problem by ordering the binary codes such that successive codes differ in only a single bit (Bitner et al., 1976; Scharstein and Szeliski, 2003). This ensures that adjacent tiles have bit patterns that differ in only one position. Camera beams that strike a boundary between two tiles will measure the superposition of two very similar codes, and the most likely error to occur (in the bit position that differs between the two tiles) will result in a localization error of at most one tile. In general though, the superposition of binary codes separated by large Hamming distances leads to measurements that are difficult to interpret and that lack a reliability metric.

The Bloom filter is an extremely space-efficient data structure for probabilistic set membership testing (Bloom, 1970). It is represented as a vector of $m$ bits, all initialized to 0. To insert an object, one computes $k$ independent hash values, all in the range $[1,m]$ and sets the corresponding bits to 1. To query whether an object is in the set, one computes its hash values and checks whether those bits are all on (an $O(1)$ operation). False negatives are impossible (assuming no error in reading the bit values), although there is a probability of approximately

$$f = \left(1 - \left(1 - \frac{1}{m}\right)^{kn}\right)^k \tag{1}$$

of returning a false positive, when the set contains $n$ elements. This probability is minimized by choosing $k = \lfloor (m/n)\ln 2 \rfloor$ to arrive at a false positive rate of approximately $f = (0.6185)^{m/n}$ (Kirsch and Mitzenmacher, 2006).

In the context of our pixel (tile) correspondences, the Bloom filter is constructed optically. We decide beforehand on an acceptable error rate $f$ or else a fixed image acquisition budget $m$, and compute the optimal $k$ value. Each tile is then assigned a binary code based on those $k$ uniformly-distributed hash values. Because the number of tiles is smaller than the universe of $\binom{m}{k}$ keys, it is feasible to explicitly enumerate them all as the columns of a "code matrix" $C$, as depicted in Figure 2.

The camera acquires a sequence of images, which are then thresholded to binary values. Each pixel therefore records a signal vector $y$ that corresponds to a Bloom filter containing the hash codes of all the tiles struck by that camera beam. By our assumption of near-Dirac PSFs, there is an upper bound of 4 on the number of elements in the set ($n$). With 64 images, this gives a false probability rate of approximately 0.05%. Since they are sparsely distributed, these errors can be detected via a spatial median filter.

Decoding the measured signals $y_i$ is a matter of inverting the Bloom filter. Since we have the matrix $C$, this can be done by solving the equation

$$y_i = (Cx > 0). \tag{2}$$

The underdetermined system can only be solved by assuming that $x$ is sparse, which is the case for near-Dirac PSFs. This is similar to the standard basis pursuit problem

$$\min_x ||x||_1 \text{ subject to } y_i = Cx \tag{3}$$

encountered in compressed sensing problems. Having chosen the columns of $C$ independently to be sparse binary vectors, they are incoherent (mutually orthogonal), satisfying the restricted isometry property (Candes and Tao, 2005). The primary difference between Equation 2 and basis pursuit is that we cannot measure $Cx$ directly and must make do with only its sparsity pattern. In practice, solutions can be found with the aid of heuristics. To solve the equation we first compute

$$v_i = C^T y_i. \tag{4}$$

Since the matrices are sparse and binary, this can be done for each pixel $y_i$ reasonably efficiently. The result is an integer-valued vector $v_i$. The indices of entries of $v_i$ equal to $k$ correspond to a superset of codes that make up the solution. Extracting only those columns of $C$ to form $C'$ allows us to instead solve the much smaller problem

$$y_i = C'x > 0. \tag{5}$$

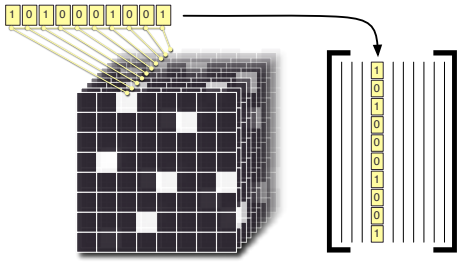


Figure 2: Binary temporal codes. Each tile is assigned a unique binary code across the projected image sequence. The codes form the columns of the code matrix.

Due to partial overlap it is possible for codes to be erroneously included in $C'$. For example, given binary codes $U = (0,1,1)$, $V = (1,1,0)$ and $W = (1,0,1)$ then a ray striking tiles coded by $U$ and $V$ will produce the measurement $X = (1,1,1)$. $W$ will then be included in $C'$ since $W \cdot X = 2 = k$. Our objective therefore is to find a minimal subset of the active codes that adequately explain the measurement.

Any algorithm for solving the basis pursuit problem will give us an estimate of the solution. We additionally impose the constraint that $0 \leq x \leq 1$. Unfortunately, since overlapping nonzeros in the codes produce values that exceed the range of $y_i$, an exact solution is unlikely and we must instead threshold the resultant $x$ at an empirically-determined value (0.1 in our experiments).

Another heuristic is to enforce spatial coherence, which will be satisfied by all near-Dirac beams. The tile coordinates corresponding to the codes in $C'$ are clustered according to a Chebychev distance threshold of 1. This gives us separate islands of tiles, each of which is checked to see if its constituent tile codes can account for all the observed "on" pixels. If so, then that one island is a solution to Equation 5.

During processing, any pixels that cannot be decoded are recorded for further examination during the postprocessing phase. At that time, the neighbors have been determined, so any tile islands that lie suffiently close to any of the neighbors are considered to be valid solutions, even if a few code bits do not match (the result of thresholding errors during acquisition).

## 3.3 Intra-tile Coding

When a camera beam neither splits into multiple paths, nor spreads out over a large area, we expect a small PSF lying either entirely within one tile, or across the boundaries of two, three or four neighboring tiles. Because the pixels struck by a beam within a tile are somewhat analogous to the tiles struck on

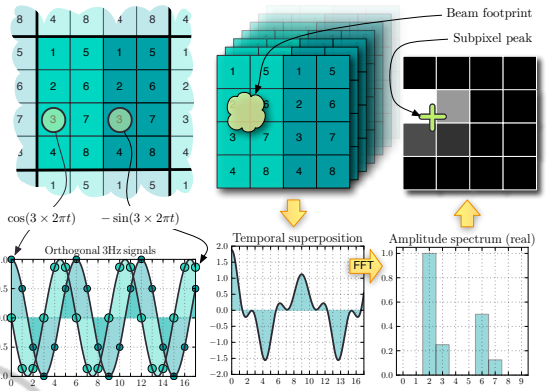



Figure 3: Left: Sample 4x4 tile. In this case, $f_{\max} = 8Hz$. The phase is 0 in the left half of the tile and $\pi/2$ in the right. Right: Temporal superposition of signals under the beam footprint.

the background, we could use the same strategy for detecting them: uniquely coding each pixel within a tile. There are however, two key differences here that call for a different method. First, a greater proportion of the pixels within a tile will be struck than the proportion of tiles struck within the background. Both inter-tile coding methods break down when too many codes are superimposed. Second, there are relatively few pixels in total within each tile, making a more direct, non-parametric method feasible.

In particular, we use the frequency coding method described by Zhu and Yang (2004), but modify it to require only half as many images. Figure 3 shows an example of a tile surrounded by segments of its eight neighbors. The $N \times N$ tile is split vertically in half, and each side is enumerated as indicated. The label of the $k$'th pixel corresponds directly to its temporal frequency $f_k$. The spatial location is encoded by setting the frequency and phase of a complex exponential

$$s(t) = Ae^{i(2\pi f t + \phi)}, \quad t \in [0,1) \tag{6}$$

and modulating the intensity $s$ of each background pixel over time as

$$s'(t) = \lfloor 0.8D/2 \rfloor s(t) + D/2 \tag{7}$$

in order to take the effective dynamic range $D$ of the display into account ($D = 256$ for an 8-bit LCD). The factor 0.8 is chosen empirically to avoid the extremes of the display's intensity range, where clipping can occur. The pixel's location is hence transmitted as a sampled waveform. The maximum frequency $f_{\max}$ is $N^2/2$Hz and so we set $F_s = 2(f_{\max} + 1)$ to satisfy the Nyquist rate, and the sampling rate to $T_s = 1/F_s$. The projected frames then correspond to discrete times $n \in \{0, T_s, 2T_s, \ldots, 1 - T_s\}$. If the phase were unrestricted we could generate a discrete sequence of pixel intensities for the $k$'th pixel as

$$s'_k[n] = s'(s_k[n]) = s'\left(Ae^{i(2\pi f_k t+\phi_k)}\right). \quad (8)$$

But we can ease the spectral estimation by allowing only two phases spaced exactly one quarter-period apart, chosen for convenience to be 0 and $\pi/2$. Hence we assign the following signals within a tile:

$$s_k[n] = \begin{cases} \cos(2\pi f_k n), & \text{left half,} \\ -\sin(2\pi f_k n), & \text{right half.} \end{cases} \quad (9)$$

The encoder assigns to all signals a unit magnitude ($A = 1$) so that we can easily recover relative contributions from multiple frequencies when camera rays strike multiple pixels. The receiver measures a superposition of signals from the $p$ pixels struck by the beam, corrupted by what we model as additive white Gaussian noise $w$:

$$x[n] = \sum_{l=1}^{p} A_l e^{i(2\pi f_l n+\phi_l)} + w[n] \quad (10)$$

Our goal is to estimate the parameters $f_l$ and $\phi_l$, which together encode the position of each component pixel, and $A_l$ which will represent the relative amount of light arriving at the sensor from it. To estimate these spectral parameters we use the periodogram, which represents the magnitude-squared Fourier transform of the signal, divided by the number of time samples (Kay, 1993). After performing a per-pixel FFT, we scale by $T_s$ and discard the redundant copy of the spectrum. The real component (in-phase channel) then directly corresponds to the relative contribution $A_k$ towards the PSF from pixels in the left half and the imaginary component (quadrature channel) likewise corresponds to contributions from the right. The PSF can be directly visualized by plotting these results as an NxN intensity plot, as in Figure 3 (right, top). It is thus described non-parametrically, and a subpixel-accurate location of the peak may be interpolated and added to the tile's global coordinates. An approximate interpolant may be obtained via the amplitude spectrum's centroid, or a local $3 \times 3$ Gaussian fit (Thomas et al., 2005).

A complication arises if the beam crosses a tile boundary. Previous methods for handling boundary overlaps in tile-based schemes have involved scanning additional passes with translationally offset tile grids (Sen et al., 2005) and considering only one of these passes: that which finds the PSF fully enclosed by a single tile. Our method requires only a single pass, as long as the PSF is smaller than a single tile. We locate the maximum value in the magnitude spectrum and circularly shift this to the centre of the tile, recording the shift vector so that we can subtract it and still obtain an absolute position in global coordinates.

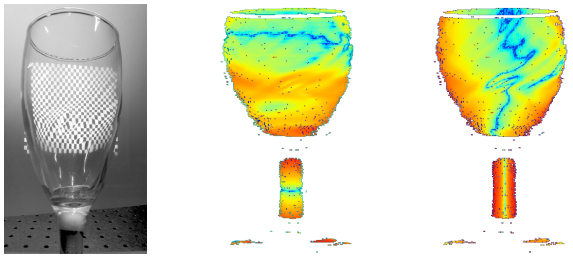


Figure 4: Background distortion through a poorly-manufactured wineglass. The rightmost images show log magnitude of vertical and horizontal apparent displacement of background pixels when viewed through the glass. The scale is 0.01 to 250 pixels.

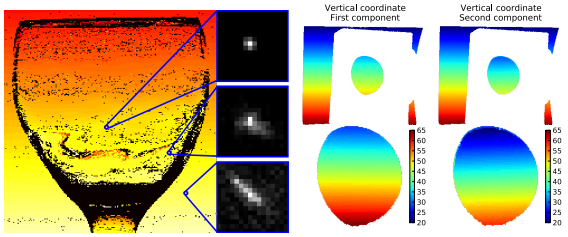
## 4 RESULTS

We first demonstrate successful capture of simple environment mattes using the binary/frequency coding method, then present simulations indicating the expected performance of a method based on high resolution spectral estimation. We include them since they suggest a way to increase the information throughput for a given image aquisition budget, but note that a more accurate measurement setup would be necessary to achieve such results in practice.

To test the algorithm we computed optical flow by comparing the correspondences before and after placing a refracting object in front of the camera. Figure 4 shows the displacement vectors and a sample photograph of the scene (from a different viewpoint). The Bloom filter parameters for this dataset were $m = 60$ and $k = 4$. Aside from missing data in regions of total internal reflection, the errors are few and easily filtered out.

In some cases we require a single corresponding point on the background for each camera pixel, in others we require the whole PSF. Figure 5(a) shows how our method can provide both an accurate non-parametric PSF, as well as a reasonably accurate point correspondence. Since the precise location of non-Dirac PSFs is undefined, we choose it to be the centroid of the neighborhood around the brightest pixel. In a moving scene one may compute the optical flow between PSFs from one time step to the next, without needing to know their precise location. Figure 5(a) also shows a near-failure case where too few binary code images were captured ($m = 40$, $k = 4$), resulting in many undetectable pixels.

Unlike Gray codes, our method is capable of detecting PSFs composed of multiple near-Dirac components. Figure 5(b) shows an example where a beamsplitter (mounted inside the occluding housing) and mirror combination are used to direct camera




(a) (b)

Figure 5: (a) Examples of spread-out, bimodal and point-like point spread functions. The color gradient indicates the vertical component of the detected background pixels' coordinates. (b) Multipath correspondences. A beamsplitter inside the occluder reflects some light onto a mirror that directs it to another point on the illuminant. The bottom row of images show a closeup of the central region (the beamsplitter) on a different color scale.

rays to two distinct points on the illuminant. This example shows only the inter-tile binary coding result, since frequency-based intra-tile coding would require larger tiles when acquiring larger, or multi-component, PSFs. In this case, we eliminate the tiles and apply binary coding to each pixel. The result is that fewer images need be captured, at the cost of losing subpixel precision.

Capture paramaters were $m = 112$ and $k = 10$. We assumed that at most 8 tile codes would be present in any one Bloom filter to accommodate the worst case of both beams striking at the intersection of four neighboring tiles. The false positive probability in this case is 0.098%.

To verify the accuracy of our frequency estimations and to determine appropriate parameter values, we conducted simulations under expected conditions. Figure 6 shows the results. In the leftmost graph, we analyzed the impact of measurement noise for the case where only a single frequency is embedded in the signal. The graphs show median absolute error, relative to the Nyquist frequency, so an upper error value of $0.5 \times 10^{-3}$ indicates that we could choose a sampling lattice spacing of double this, i.e. $\delta_f = 0.001 \times N/2$ Hz. Error values asymptotically approach a lower bound as the number of captured images increases, but going beyond 100 images leads to diminishing returns. Too few images however, lead to very high error, indicating that ESPRIT would not be suitable for detecting frequencies within a tile.

Next, we investigated how superposition of signals degrades performance. The second graph shows cases with up to four simultaneous frequencies, chosen randomly, but spaced far enough apart so as not to be strongly correlated. The amplitudes were all set to 1.0 and the simulation was run at an SNR of 30dB. Accuracy does degrade as more signals are added, but the effect becomes relatively small as $N$ increases.

The third graph shows that we are unable to detect phase as accurately as frequency. For this reason, the sampling interval $\delta_\phi$ depicted in Figure 1 must be much larger than $\delta_f$. The vertical axis in this graph is relative to $\pi$ rad/sample.

The final graph shows amplitude accuracy, at which we obtain similar performance to phase (as is to be expected, since both values result from the solution of the same linear system). The vertical axis is relative to the unit input signal amplitude.

Given these results, we can determine the number of tiles than can adequately be coded given a fixed image acquisition budget. For a typical case of $N = 64$ images taken at an SNR of 30dB, when four sinusoids are present, frequency can reliably be detected to within $0.0004 \times N/2 = 0.0128$Hz, and the phase is accurate to within $0.005 \times \pi$ rad/sample. Avoiding the lower 5% and upper 5% of frequencies, and covering this space with a lattice of points spaced $\delta_f = 2 \times 0.0128$ Hz and $\delta_\phi = 2 \times 0.005\pi$ rad/sample apart gives us 225k sample points, i.e. 64 images is enough to support 225k tiles, so long as no more than four of them are superimposed at one pixel.

## 5 CONCLUSIONS

Most prior methods for establishing pixel correspondences are based on matching spatio-temporal intensity patterns. These produce qualitatively good visual results, but lack guarantees on correctness. We have proposed instead to assign unique codes on the tile level and then demultiplex them after transmission through the optical projector-camera multiplexer. This opens up the possibility of using tools from digital signal processing to ensure that each code is accurately read. One possible direction for future work would be to insert error detection and correction codes into the signals.

Our current binary signal decoding scheme employs compressed sensing and spatial heuristics to demultiplex signals. We have introduced the Bloom filter as an optical computing tool for determining one-to-few pixel correspondences. However, without more advanced DSP techniques, we cannot accommodate one-to-many correspondences. To circumvent this problem, we group pixels into tiles, and apply a separate frequency-based coding scheme to map the pixels within each tile. To this end, we have improved upon existing methods by halving the required number of images, eliminating redundant sweep scans, and allowing for subpixel precision with nonparametric point-spread functions.

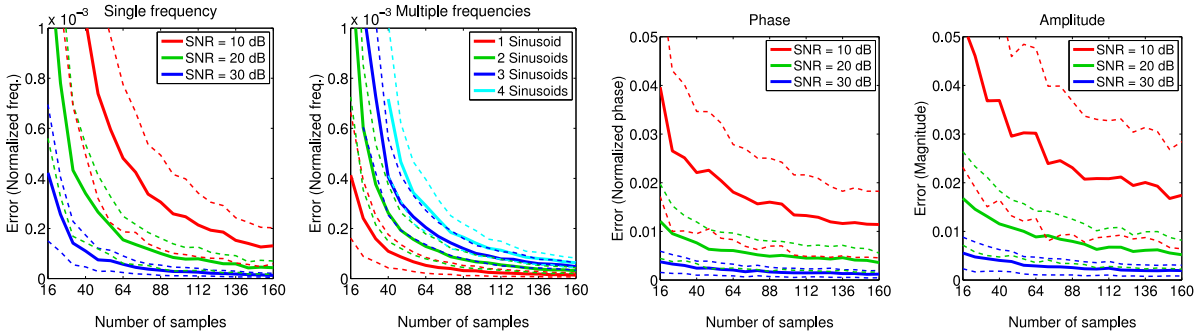



Figure 6: Synthetic experiment results. Solid lines show the median absolute error, while dashed lines indicate the median absolute deviation. 500 trials were performed for each tested sample size.

# REFERENCES


Atcheson, B., Ihrke, I., Heidrich, W., Tevs, A., Bradley, D., Magnor, M., and Seidel, H.-P. (2008). Time-resolved 3D capture of non-stationary gas flows. *ACM Trans. Graphics*, 27:132.

Bitner, J., Ehrlich, G., Reingold, E., and 1976 (1976). Efficient generation of the binary reflected Gray code and its applications. *Communications of the ACM*, 19:517–521.

Bloom, B. H. (1970). Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13:422–426.

Candes, E. J. and Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51:4203–4215.

Chuang, Y.-Y., Zongker, D. E., Hindorff, J., Curless, B., Salesin, D. H., and Szeliski, R. (2000). Environment matting extensions: Towards higher accuracy and real-time capture. In *Intl Conf. on Computer Graphics and Interactive Techniques*, pages 121–130.

Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M. (2000). Acquiring the reflectance field of a human face. In *SIGGRAPH '00*, pages 145–156.

Hall-Holt, O. and Rusinkiewicz, S. (2001). Stripe boundary codes for real-time structured-light range scanning of moving objects. In *Proc. ICCV*, volume 2, pages 359–366.

Kay, S. M. (1993). *Fundamentals of statistical signal processing, Volume I: Estimation theory*. Prentice Hall PTR.

Kirsch, A. and Mitzenmacher, M. (2006). Less hashing, same performance: Building a better bloom filter. In *Algorithms - ESA 2006*, volume 4168, pages 456–467. Springer.

Peers, P. and Dutré, P. (2005). Inferring reflectance functions from wavelet noise. In Bala, K. and Dutré, P., editors, *Proc. EGSR*, page 173182.

Roy, R. and Kailath, T. (1989). ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 37:984–995.

Rusinkiewicz, S., Hall-Holt, O., and Levoy, M. (2002). Real-time 3D model acquisition. *ACM Trans. Graphics*, 21:438–446.

Scharstein, D. and Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. In *Proc. CVPR*, volume 1, pages I–195—-I–202.

Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M., and Lensch, H. P. A. (2005). Dual photography. *ACM Trans. Graphics*, 24:745–755.

Thomas, M., Misra, S., Kambhamettu, C., and Kirby, J. (2005). A robust motion estimation algorithm for piv. *Measurement Science and Technology*, 16:865–877.

Wexler, Y., Fitzgibbon, A., and Zisserman, A. (2002). Image-based environment matting. In *Proc. 13th Eurographics Workshop on Rendering*, pages 279–290.

Zhu, J. and Yang, Y.-H. (2004). Frequency-based environment matting. In *Proc. Pacific Graphics*, pages 402–410.

Zongker, D. E., Werner, D. M., Curless, B., and Salesin, D. H. (1999). Environment matting and compositing. In *Intl Conf. Computer Graphics and Interactive Techniques*, pages 205–214.