The Minimalist Camera

Parita Pooj[1]
parita@cs.columbia.edu
Michael Grossberg[2]
http://www-cs.ccny.cuny.edu/~grossberg/
Peter Belhumeur[1]
https://www.peterbelhumeur.com/
Shree Nayar[1]
http://www.cs.columbia.edu/~nayar/

[1] Columbia University
New York, USA
[2] City University of New York
New York, USA

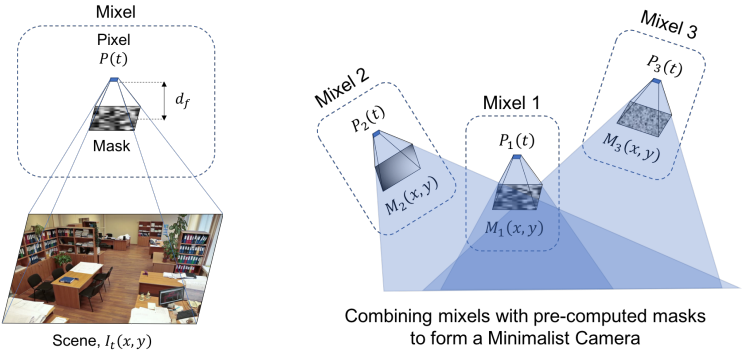
Abstract

We present the *minimalist camera (mincam)*, a design framework to capture the scene information with minimal resources and without constructing an image. The basic sensing unit of a mincam is a 'mixel' — an optical photo-detector that aggregates light from the entire scene linearly modulated by a static mask. We precompute a set of masks for a configuration of few mixels, such that they retain minimal information relevant to a task. We show how tasks such as tracking moving objects or determining a vehicle's speed can be accomplished with a handful of mixels as opposed to the more than a million pixels used in traditional photography. Since mincams are passive, compact, low powered and inexpensive, they can potentially find applications in a broad range of scenarios.

1 Introduction

Cameras were originally invented to capture images of a scene. Today we use them to capture a wide range of visual information that can then be used to automate intelligent systems such as robots. This has led to the increasing presence of cameras everywhere — in private and public spaces for security, on drones and robots for mapping the physical world, in wildlife sanctuaries for monitoring varied species, and in various industries to address domain-specific problems. In many applications, the end goal is to acquire specific information that can often be retrieved without capturing an image as the intermediate step. To this end, we present the *minimalist camera (mincam)*, a sensing architecture whose goal is to capture information needed to perform specific computer vision tasks using minimal sensing resources, without explicitly capturing an image of the scene.

With the mincam, we take a step toward answering the question: what is the minimum number of pixels needed to retrieve information relevant to a task? To answer this question, we propose a sensing architecture consisting of 'mixels', short for masked pixels. This architecture is influenced by a single pixel camera [25]; the key difference being that mixels have a *static* gray-scale mask that the scene is imaged through. Such masks are compact, light-weight, inexpensive, and do not need external power, making them practical. By carefully designing the masks, we are able to capture scene information relevant to performing a vision task using a very small number of mixels. As an example, we demonstrate that a new



(a) (b)

Figure 1: (a) A 'mixel' (masked pixel): A pixel measures a brightness of $P(t)$ at time t by aggregating light incident through the static mask, $M(x,y)$ placed at a distance of d_f. (b) A mincam combines a small number of mixels by designing masks for a specific task.

object entering the scene can be tracked using four masked pixels which is far fewer than the more than a million pixels used in a conventional camera. The proposed mincam framework takes first steps towards building a ubiquitous camera; that is, it is extremely small, simple to build and needs very low power to capture and transmit few bytes of information.

2 Related Work

The mincam builds on three areas of research: single pixel imaging, coded aperture imaging and feature-specific imaging. The architecture of single pixel imaging systems, the aperture design principle of coded aperture imaging systems and the applications of feature specific imaging systems influence the design of the mincam.

Single pixel imaging systems attempt to reconstruct the visual appearance of scenes using one or a very small number of pixels. The single pixel camera introduced by Takhar et al. [25] and Duarte et al. [8] captures multiple projections of the scene using a digital micromirror device (DMD) between the scene and a single photo-diode. More recent single pixel imaging systems capture the projections using active illumination [12, 29] and other spatial light modulators to reconstruct either a 2D image of the scene [6, 11, 20], a colored image of the scene [28] or a 3D mapping of the scene [23, 24]. Neifeld and Ke [15] provide a basic comparison between several theoretical optical systems that could be used to obtain the linear projections for minimizing the reconstruction error. The mincam differs from single pixel imaging systems in that it only attempts to capture the information relevant to performing particular vision tasks without attempting to reconstruct the scene.

Coded aperture imaging techniques make use of patterned masks as apertures to preserve useful information about the scene. They were introduced in the field of high energy astronomy to improve signal to noise ratio in lensless imaging systems [5, 9, 21] and in optics to increase depth of field [13]. In more recent works, Veeraraghavan et al. [27] recovered the 4D light field of a 2D scene using patterned masks and Raskar et al. [19] achieved motion deblurring using a coded exposure camera. The aperture designs used for both these systems aim to preserve high spatial frequencies using patterned masks. Zhou and Nayar [30] devise an optimization criteria to find high resolution coded apertures for defocus deblurring. This is later leveraged by Zhou et al. [31] to estimate depth from defocus. While these coded apertures are designed to be used with conventional cameras, the design of coded masks for

a mincam are optimized for a lensless imaging system.

Feature-specific imaging systems capture image features directly to perform particular vision tasks. Neifeld and Shankar [16], for example, propose a polarization-based optical system to capture linear object features such as Hadamard and wavelet features directly. Ashok et al. [3] perform target detection by obtaining features from several projection bases such as principal components optimized by the task-specific information (TSI) metric [17]. (They later design TSI-optimized projections for improving the reconstruction of the scene [4].) Davenport et al. [7] classify and recognize objects by using maximum likelihood hypotheses with random projection features. Stenner et al. [22] and Townsend et al. [26] track objects simulated as dots moving on a screen using compressive sensing features.

Like other feature-specific imaging systems, the mincam attempts to extract features for performing specific computer vision tasks. Unlike these systems, however, the mincam is passive; it does not use any active elements such as DMDs, active illumination or spatial light modulators. To the best of our knowledge, our design framework provides the most simple and practical implementation of the basic idea of feature-specific imaging. Our framework is also powerful in that it can be extended: we demonstrate designs for detecting intrusion, tracking moving objects and estimating the speed of moving objects, but our framework can be extended to many other applications.

3 The Minimalist Camera

A mincam consists of one or more masked pixels (mixels), as shown in Figure 1. Each mixel consists of an optical photo-detector (or a pixel) that looks through a static mask with two-dimensional intensity. For example, the mask could be implemented as a fixed grayscale pattern on a transparent film. Each mixel measures linear projection of incident scene radiance $I(x,y)$, where $I(x,y)$ is defined as the 2D image that would have been captured by a conventional pinhole camera in the same location as the mixel. An example is demonstrated in Figure 1(a). At time t, scene radiance represented as an $M \times N$ matrix $I_t(x,y)$ is multiplied by the mask $M(x,y)$ and is aggregated into a measurement $P(t)$ where,

$$P(t) = \int_0^N \int_0^M M(x,y) I_t(x,y)\, \mathrm{d}x\, \mathrm{d}y. \tag{1}$$

The mincam, consisting of multiple mixels, can thus be defined as a camera that reduces the scene radiance to a lower dimension using linear projections. The number of required linear projections determine the number of mixels in a mincam and the choice of masks is determined by the linear projections useful for a given application.

3.1 Hardware Architecture

A mixel is an optical photo-detector with a high-resolution mask printed on transparent material. As a proof of concept, we emulate this using off-the-shelf components. We use a Ximea Subminiature camera for constructing each mixel. Using a commercial printer, we print each mixel's mask on a transparent film. We then place the mask at the mixel's aperture so that it linearly modulates the light before it reaches the mixel's sensor. The 48×48 px region of interest at the center of the sensor array is averaged to obtain the mixel measurement. The effective mixel, thus, occupies an approximately $0.1\,\mathrm{mm} \times 0.1\,\mathrm{mm}$ area of the sensor array and the measurement is quantized to 12 bits.

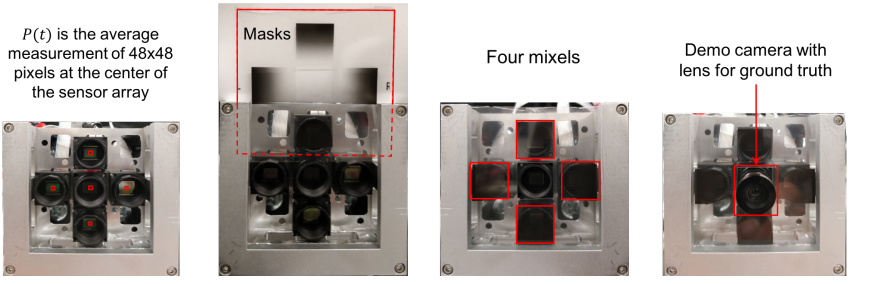


Figure 2: Hardware architecture: (a) The grid of 48×48 pixels at the center of a camera's sensor array is averaged to obtain one 12-bit mixel measurement, $P(t)$. A subminiature camera's sensor array is used for each mixel. (b) The mask design for four outer mixels is slid into the slot. (c) Four outer cameras become mixels after the mask is placed. (d) An example of a mincam prototype with four mixels and a center camera with lens to capture ground truth.

We use different configurations of mixels to design mincams for various applications. Figure 2(a) shows a configuration of five Ximea Subminiature cameras placed closely together. As 2(b) shows, we create mixels from these cameras by sliding film masks into a slot in front of the cameras. For some applications, we attach a lens to the center camera as shown in Figure 2(d) to capture ground truth images in order to demonstrate our results.

Based on the hardware requirements, a mixel can be manufactured with a size of less than 1 mm×1 mm×1 mm with a high-resolution mask placed very close to the photo-detector. Every mixel captures 12 bits of information that can be transmitted to a processing hub with high frame rate and low power. The mask is optimally light efficient for capturing a projection since it collects maximum ambient light from the scene and its static design allows the mincams to be manufactured as tiny, low powered devices. high dynamic range measurements. A mincam which is a combination of very few such mixels can thus be considered as a small, inexpensive, low-powered device that can potentially become self-powered in the future [1, 2, 14].

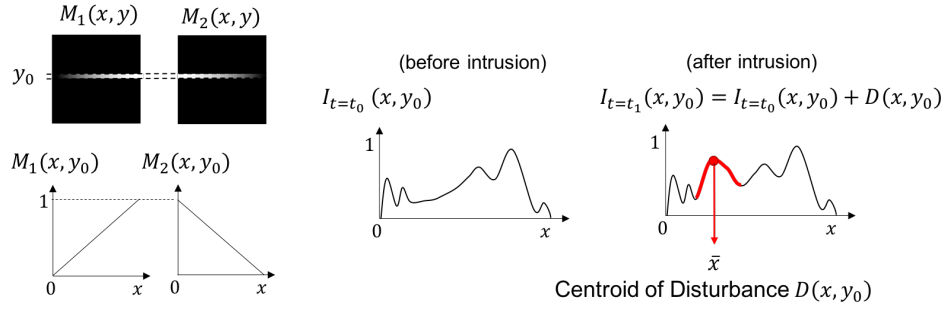
4 Example Applications of a Minimalist Camera

The applications demonstrated in this section show how a mincam design framework can be used to apply towards any application with a focus on minimalism and simplicity. In this section, we show how a mincam can locate an intruding object along a boundary, track a moving object and determine the speed of a moving vehicle using four or fewer mixels.

4.1 1D Intrusion Detection

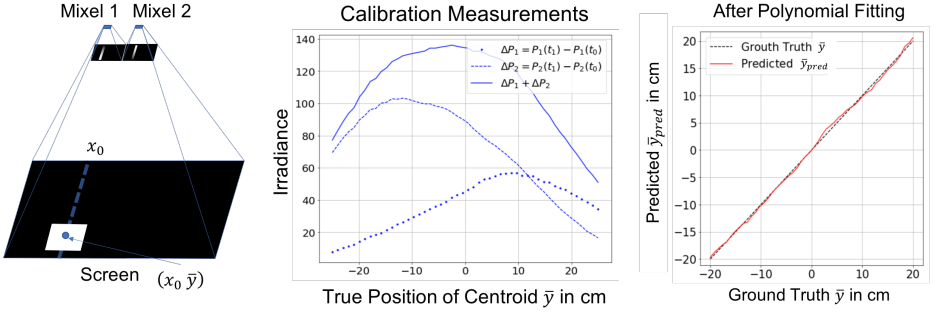
We design a mincam with two mixels to detect intrusion along a virtual boundary. This boundary can be defined along a 1D line or a 1D curve in a scene. The finish line of a car race is an example of such a virtual boundary. To detect intrusion, we are interested in determining when and where it occurred on the boundary. Traditional computer vision techniques convert the scene to a binary image and then compute the centroid of the intruding object to locate it. To simplify the problem, we assume that the centroid of the intensity change caused by an intruding object is approximately the location of intrusion.

We prove that two mixels with masks shown in Figure 3(a) are sufficient to obtain this



(a) (b)

Figure 3: 1D intrusion detection at boundary, $y = y_0$ (a) Masks designed for the two mixels to detect and locate intrusion. (b) The intruding object changes the initial intensity, $I_{t=0}(x,y_0)$ by the disturbance, $D(x,y_0)$. We recover the centroid of this disturbance, $\bar{x}$.



(a) (b) (c)

Figure 4: Calibration for 1D intrusion detection at boundary, $x = x_0$ (a) We display a white square as disturbance along the dotted boundary on the screen for calibration. (b) The figure plots the difference of mixel measurements in the presence and absence of the white square against each position at which the white square is displayed. Physical limitations do not allow the mixels to view the same scene leading to asymmetry in the measurement of $\Delta P_1 + \Delta P_2$. (c) The observed centroid position are fitted to ground truth using polynomial fitting.

information. Both mixels are assumed to have the same field of view. This assumption is reasonable since it can be approximated with a tiny mixel of size less than 1mm×1mm×1mm. Without loss of generalization, the boundary is considered along a straight line at $y = y_0$ for readability. In mathematical form, the masks can be written as

$$M_1(x,y_0) = x, \qquad M_2(x,y_0) = 1 - x. \tag{2}$$

Figure 3(b) illustrates how the intensity along the boundary changes by a disturbance $D(x,y_0)$ at time $t = t_1$ when an object intrudes at the boundary. We make use of this temporal change between the intensity before intrusion at time t_0 and after the intrusion at time t_1. The change in mixel measurements as given from Equation (1) are thus equivalent to

$$P_1(t_1) - P_0(t_0) = \int x D(x,y_0)\,\mathrm{d}x, \qquad P_2(t_1) - P_2(t_0) = \int (1-x) D(x,y_0)\,\mathrm{d}x. \tag{3}$$

The limits for the integral are eliminated for the sake of simplicity and assumed to be over the

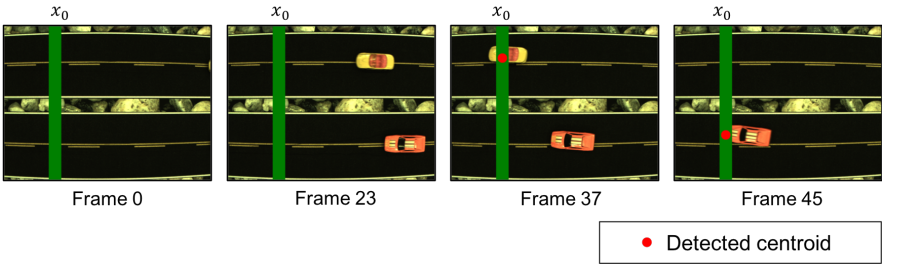


Figure 5: Results of the physical experiment conducted using two toy cars for demonstrating the application. The cars crossing the finish line boundary at $x = x_0$ are detected and located at the red dot for the position of intrusion. The correct position is predicted with a maximum error rate of 11.42% and an average error rate of 4.83%. A complete video of the car race is provided in the supplementary material for proper analysis of results.

field of view of the mixel. When the temporal change in measurements given by Equation (3) is above a threshold, we get the centroid of the detected disturbance,

$$\bar{x} = \frac{\int xD(x,y_0)\,\mathrm{d}x}{\int D(x,y_0)\,\mathrm{d}x} = \frac{P_1(t_1) - P_1(t_0)}{P_1(t_1) - P_1(t_0) + P_2(t_1) - P_2(t_0)}. \tag{4}$$

This result can be generalized for any boundary along a 1D curve defined by $s = f(x,y)$. For example, a circular boundary or a closed curve around an object of interest can be considered a virtual boundary. For these cases, the linear variation of the mask will be along the circumference of the circle or along the curved boundary s.

We reiterate that we find the centroid of the intensity change caused by the object and not the centroid of the object. For instance, an object with uneven brightness will be located with a slight bias in favor of the brighter part and if two objects intrude the boundary at the same time, the location of intrusion will lie between them signifying the centroid of the intensity change collectively caused by both. Despite these conditions, we now show that our system measures the approximate location of intrusion with satisfactory precision.

4.1.1 Experiments: Calibration and Results

To demonstrate how a two-mixel mincam can perform intrusion detection empirically, we use two cameras for mixels and one camera to capture ground truth video. We perform a one-time offline calibration using a 50" screen in the dark to simulate a scene. The ground truth is controlled by programming the display.

For this example experiment, the boundary is the finish line at $x_0 = -13.78$ cm as shown by the green line in Figure 5. We calibrate the system by displaying a white square on the screen along this boundary as shown in Figure 4(a). We visualize the difference in mixel measurements represented by Equation (3) in Figure 4(b). Equation (4) is applied to the mixel measurements obtained from calibration to calculate intrusion position. The observed intrusion position, $\bar{y}_{obs}$ is a relative value between 0 and 1. We expect the position to vary linearly as the white object intrudes at uniform intervals from $y = -25$ cm to $y = 25$ cm. However, since our mixels are placed 3.3 cm apart, we observe that the shift in their respective fields of view introduces minor non-linearities. We represent the predicted position as a function of $\bar{y}_{obs}$, $\bar{y}_{pred} = f(\bar{y}_{obs})$ where $f(.)$ is assumed to be a 3^{rd} degree polynomial to account for non-linearities introduced by drawbacks of hardware realization.

The ground truth position known during calibration is used to compute the function $f(.)$ which is later applied directly to data acquired from real scenes when using real-life objects for intrusion to predict the correct centroid of intrusion.

To evaluate the system, we race two toy cars on a printed background. For the better understanding of the results, the complete video of the experiment is provided in the supplementary material. Some example frames from the video are demonstrated in Figure 5. As is clear from the results, the location of a car crossing the boundary is satisfactorily identified in this experiment.

4.2 Naïve Object Tracking

Extending the intuition used in Section 4.1, we design a mincam to detect and track a new object introduced in the scene. Four mixels with masks as illustrated in Figure 6 are found to successfully track any new object moving on a static background. We repeat the assumption from Section 4.1 here by claiming that the location of the object being tracked can be approximated by the centroid of the change in intensity it causes. Thus, the masks are a 2D extension of the previous application and are formulated as

$$\begin{aligned} M_1(x,y) &= x, \quad M_2(x,y) &&= 1-x, \\ M_3(x,y) &= y, \quad M_4(x,y) &&= 1-y. \end{aligned} \tag{5}$$

For each mixel, the measurement at any time $t = t_1$ when the object is visible is compared to the initial measurement for the background at $t = t_0$. The difference in measurement is,

$$\begin{aligned} P_1(t_1) - P_1(t_0) &= \int\int xD(x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad P_2(t_1) - P_2(t_0) &&= \int\int (1-x)D(x,y)\,\mathrm{d}x\,\mathrm{d}y, \\ P_3(t_1) - P_3(t_0) &= \int\int yD(x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad P_4(t_1) - P_4(t_0) &&= \int\int (1-y)D(x,y)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned} \tag{6}$$

When this difference is above a threshold, we detect the presence of the object and locate it by finding the centroid of the change it causes:

$$\begin{aligned} \bar{x} &= \frac{\int\int xD(x,y)\,\mathrm{d}x\,\mathrm{d}y}{\int\int D(x,y)\,\mathrm{d}x\,\mathrm{d}y} = \frac{P_1(t_1) - P_1(t_0)}{P_1(t_1) - P_1(t_0) + P_2(t_1) - P_2(t_0)}, \\ \bar{y} &= \frac{\int\int yD(x,y)\,\mathrm{d}x\,\mathrm{d}y}{\int\int D(x,y)\,\mathrm{d}x\,\mathrm{d}y} = \frac{P_3(t_1) - P_3(t_0)}{P_3(t_1) - P_3(t_0) + P_4(t_1) - P_4(t_0)}. \end{aligned} \tag{7}$$

While it can be seen that the above derivation needs just three masks designed as below:

$$M_1(x,y) = x, \qquad M_2(x,y) = y \quad \text{and,} \quad M_3(x,y) = 1. \tag{8}$$

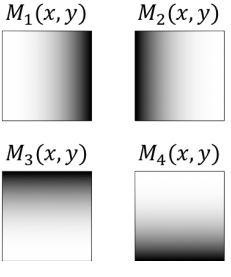


Figure 6: Mask designs for naïve object tracking

we choose the four masks designed with Equation (5) since the denominator in Equation (7) normalizes the intensities across multiple pixels and hence, is more robust.

4.2.1 Experiments: Calibration and Results

For this application, we use four cameras for the mixels and one additional camera to capture ground truth for demonstration. The configuration of Figure 2 is used with the masks printed

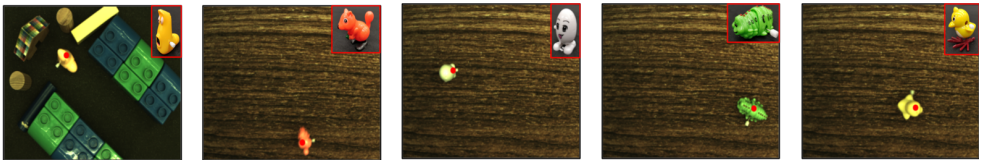


Figure 7: Results for naïve object tracking. We use wind-up toys as moving objects against a static background. Close-up shots of the tracked objects are shown in the top right corner of each image. The predicted position (red dot) is correctly identified within a maximum error rate of 6.78% with an average error rate of 2.82%. The complete video results for all toys are provided in the supplementary material.

as shown in Figure 6. For object tracking, a calibration process similar to that of 1D intrusion is used — a white square is displayed along grid points within a 50 cm diameter field of view on the screen. Using Equation (7), the relative positions are predicted and then translated to the true positions in cm using polynomial fitting. In this case, the polynomial fitting algorithm used for 1D intrusion case is extended for both x and y coordinates. We apply this calibrated fitting to measurements from real-life physical experiments. Figure 7 summarizes the results of tracking different objects moving along a static background. Each object is a wind-up toy that moves along a random path in the scene. We have provided complete videos for the results in the supplementary material along with the example frames here. We analyze our results by comparing the predicted position to the true centroid of the object obtained using background subtraction and binary object detection. Our results predict the position correctly within a maximum error rate of 6.78% with an average error rate of 2.82% measured over a range of toys as seen in the result videos.

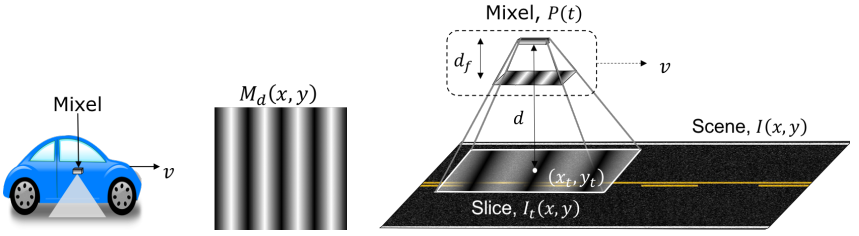
4.3 Speed Estimation

We can also use a mincam to estimate the speed of an object that it is attached to with just a single mixel. As Figure 7(a) shows, for example, a mixel attached to the side of a car facing downwards can predict its speed. There are two assumptions made in this scenario that the direction of motion is along a known axis and that the depth of the scene is known. Both of these assumptions are trivially true in the case of a car since the motion of a four-wheeled vehicle is tangential to its side and the depth of the scene that the mincam views is equivalent to the distance between the mincam and the road. Without loss of generality, we assume the direction of motion to be along x-axis and the depth of scene to be a constant, d.

For this mincam, a 2D sinusoid with a single frequency in the x-direction is the choice of mask for the only mixel, as illustrated in Figure 8(b). Zhang et al. [29] explore the choice of phase-shifted sinusoids for mask projections to sample the scene in Fourier domain. In our case, the phase shift is simulated by a vehicle's motion along the x-axis. Intuitively, the mixel samples a spatial sinusoid as it moves over the scene where the sampling rate is determined by the speed of its movement. Thus, the signal $P(t)$ is a sinusoid that encodes the velocity in its frequency, by compressing or expanding the spatial sinusoid. Based on this intuition, we show how a mixel with a sinusoidal mask can be used to measure the speed.

If the depth is considered to be $z = 0$ at the pixel, the mask is placed at $z = d_f$ and is defined with the angular frequency ω_{d_f} as

$$M_{d_f}(x,y) = A_0 + A_1 \cos(\omega_{d_f} x + \phi). \tag{9}$$



(a) (b) (c)

Figure 8: (a) A mixel attached to the side of a moving car (b) Mask for the mixel to estimate speed. (c) mixel moves over the scene, viewing a slice of the scene shown as $I_t(x,y)$.

At scene depth, $z = d$, the mask is seen with a lower frequency —

$$M_d(x,y) = A_0 + A_1 \cos(\omega_d x + \phi), \text{ where } \omega_d = \omega_{d_f} \frac{d_f}{d}. \tag{10}$$

Figure 8(c) shows an analysis of this mask moving over the scene. At time t, the mask views a slice $I_t(x,y)$ of size M×N from the scene $I(x,y)$. Thus, the mixel measurement at time t is

$$P(t) = \int_{y_t - N/2}^{y_t + N/2} \int_{x_t - M/2}^{x_t + M/2} I_t(x,y) M_d(x,y)\, dx\, dy. \tag{11}$$

For a moving vehicle, x_t changes as function of speed. As the mixel moves along the x-axis on the scene, $I_t(x,y)$ becomes a sliding window on the scene with $x_t = vt$, where v is the unknown speed of the vehicle. By extension, $P(t)$ is the horizontal slice of the correlation between $I(x,y)$ and $M_d(x,y)$ at $y = y_t$. Since the mixel's mask is a sinusoid with no vertical frequencies, Fourier slice theorem [10, 18] allows us to evaluate it in 1D along the x-axis. Correlation in spatial domain translates to multiplication in Fourier domain and the Equation (11) translates to

$$P(u) = \mathcal{F}[I_t(x,y)] \times \delta(u_x = \omega_d). \tag{12}$$

The expected signal for $P(t)$ is thus a sinusoid of frequency ω_d sampled as a function of x_t.

$$\begin{aligned} P(t) &= A_0' + A_1' \cos(\omega_d x_t + \phi'). \\ P(t) &= A_0' + A_1' \cos(\omega_d v t + \phi'). \\ P(t) &= A_0' + A_1' \cos(\omega_p t + \phi'). \end{aligned} \tag{13}$$

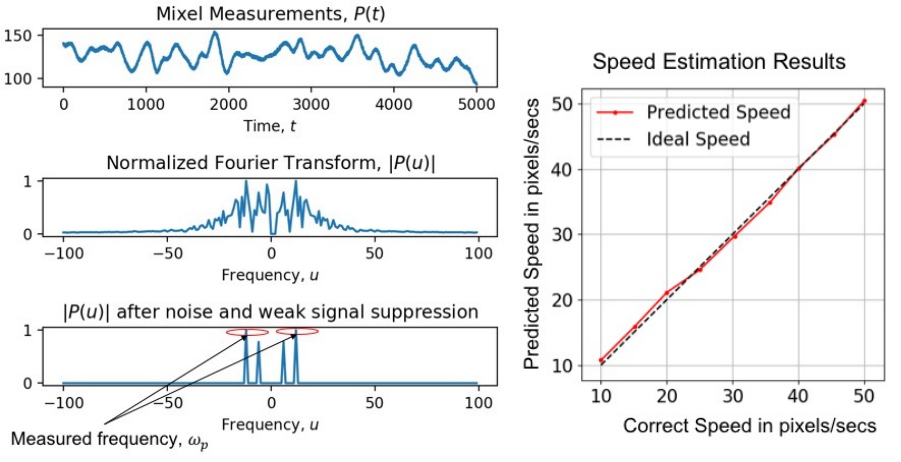
With Fourier analysis as demonstrated in Figure 9(a), the frequency $\omega_p = \omega_d v$ can be recovered. With the known ω_d, we calculate the unknown speed of the moving object to be

$$v = \frac{\omega_p}{\omega_d}. \tag{14}$$

We can extend the notion of speed estimation by using two mixels, one on each side of the car, as the mincam. With the speed estimated by each, and the known width of the vehicle, we can predict the trajectory of the moving vehicle.

4.3.1 Experimental Results

To measure the speed of a moving object, we use one camera for the mixel with a sinusoid printed on a transparent film. We simulate the ground truth by displaying a video of a moving road on a screen. Each frame in this video is a crop of the road texture shown in Figure



(a) (b)

Figure 9: (a) Example of mixel measurements as a period of time. ω_p is estimated from the measurements as depicted. (b) Speed is measured in terms of the pixels travelled by the road scene on the display in one second, as observed by the stationary mixel. The speed predicted from ω_p is plotted against the correct speed. The prediction is within a maximum error rate of 7.55 % with an average error rate of 2.85%.

8(c). We shift the crop window by one pixel resolution of the road texture image for consecutive frames. The frames per second resolution of the video displayed on the screen is thus equivalent to a shift in pixels per second, where the shift in pixels signifies the distance the moving object has covered. Thus, the desired speed is simulated in terms of the amount of shift in pixels per second. The mixel measures the intensity asynchronously to obtain the signal $P(t)$. Figure 9(a) displays one such signal $P(t)$ and the frequency ω_p extracted from its Fourier transform. We then predict the speed based on Equation (14) as plotted in Figure 9(b). We evaluate the system by noting that the maximum error rate is 7.55 % and the average error rate is 2.85%. Thus, the system predicts the speed of the vehicle accurately for our testing framework.

5 Conclusion

We provide a general framework for designing cameras that can perform vision tasks with minimal resources and without contructing an image. Our design employs static masks and can thus be manufactured as passive, low-powered devices at very low cost. It can also measure visual information in settings where capturing images might risk a breach of privacy since it is not possible to reconstruct the scene in most scenarios. Though we demonstrate mincams that perform three specific vision tasks, our framework can be applied to many more applications. We demonstrate simple applications that show that mincams could solve practical and useful tasks. The progress in information theory and linear dimensionality reduction provide a suitable direction towards designing mincams for complex applications such as object classification, depth estimation and other tasks that rely on visual information. As a concept that aims to capture useful information — and only that — with the bare minimum resources, a mincam provides a simple and practical way to re-think the purpose and, consequently, the design of a camera.

References

[1] Khondker Zakir Ahmed, Mohammad Faisal Amir, Jong Hwan Ko, and Saibal Mukhopadhyay. Reconfigurable 96× 128 active pixel sensor with 2.1 μw/mm 2 power generation and regulated multi-domain power delivery for self-powered imaging. In *European Solid-State Circuits Conference, ESSCIRC Conference 2016: 42nd*, pages 507–510. IEEE, 2016.

[2] A Anvesha, Shaojie Xu, Ningyuan Cao, Justin Romberg, and Arijit Raychowdhury. A light-powered smart camera with compressed domain gesture detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[3] Amit Ashok, Pawan K Baheti, and Mark A Neifeld. Compressive imaging system design using task-specific information. *Applied optics*, 47(25):4457–4471, 2008.

[4] Amit Ashok, Liang-Chih Huang, and Mark A Neifeld. Information optimal compressive sensing: static measurement design. *JOSA A*, 30(5):831–853, 2013.

[5] E Caroli, JB Stephen, G Di Cocco, L Natalucci, and A Spizzichino. Coded aperture imaging in x-and gamma-ray astronomy. *Space Science Reviews*, 45(3-4):349–403, 1987.

[6] Hongwei Chen, Zhiliang Weng, Yunhua Liang, Cheng Lei, Fangjian Xing, Minghua Chen, and Shizhong Xie. High speed single-pixel imaging via time domain compressive sampling. In *CLEO: Applications and Technology*, pages JTh2A–132. Optical Society of America, 2014.

[7] Mark A Davenport, Marco F Duarte, Michael B Wakin, Jason N Laska, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk. The smashed filter for compressive classification and target recognition. In *Computational Imaging V*, volume 6498, page 64980H. International Society for Optics and Photonics, 2007.

[8] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008. ISSN 1053-5888. doi: 10.1109/MSP.2007.914730.

[9] Stephen R Gottesman and EE Fenimore. New family of binary arrays for coded aperture imaging. *Applied optics*, 28(20):4344–4352, 1989.

[10] Marc Levoy. *Volume rendering using the Fourier projection-slice theorem*. Computer Systems Laboratory, Stanford University, 1992.

[11] Jianwei Ma. Single-pixel remote sensing. *IEEE Geoscience and Remote Sensing Letters*, 6(2):199–203, 2009.

[12] Filipe Magalhães, Francisco M Araújo, Miguel V Correia, Mehrdad Abolbashari, and Faramarz Farahi. Active illumination single-pixel camera based on compressive sensing. *Applied optics*, 50(4):405–414, 2011.

[13] M Mino and Y Okano. Improvement in the otf of a defocused optical system through the use of shaded apertures. *Applied Optics*, 10(10):2219–2225, 1971.

[14] Shree K Nayar, Daniel C Sims, and Mikhail Fridberg. Towards self-powered cameras. In *Computational Photography (ICCP), 2015 IEEE International Conference on*, pages 1–10. IEEE, 2015.

[15] Mark A Neifeld and Jun Ke. Optical architectures for compressive imaging. *Applied optics*, 46(22):5293–5303, 2007.

[16] Mark A Neifeld and Premchandra Shankar. Feature-specific imaging. *Applied optics*, 42(17):3379–3389, 2003.

[17] Mark A Neifeld, Amit Ashok, and Pawan K Baheti. Task-specific information for imaging system analysis. *JOSA A*, 24(12):B25–B41, 2007.

[18] Ren Ng. Fourier slice photography. In *ACM transactions on graphics (TOG)*, volume 24, pages 735–744. ACM, 2005.

[19] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *ACM Transactions on Graphics (TOG)*, 25(3): 795–804, 2006.

[20] G. Satat, M. Tancik, and R. Raskar. Lensless imaging with compressive ultrafast sensing. *IEEE Transactions on Computational Imaging*, 3(3):398–407, Sept 2017. doi: 10.1109/TCI.2017.2684624.

[21] Gerald K Skinner. X-ray imaging with coded masks. *Scientific American*, 259(2): 84–89, 1988.

[22] Michael D Stenner, Daniel J Townsend, and Michael E Gehm. Static architecture for compressive motion detection in persistent, pervasive surveillance applications. In *Imaging Systems*, page IMB2. Optical Society of America, 2010.

[23] Baoqing Sun, Matthew P Edgar, Richard Bowman, Liberty E Vittert, Stuart Welsh, A Bowman, and MJ Padgett. 3d computational imaging with single-pixel detectors. *Science*, 340(6134):844–847, 2013.

[24] Ming-Jie Sun, Matthew P Edgar, Graham M Gibson, Baoqing Sun, Neal Radwell, Robert Lamb, and Miles J Padgett. Single-pixel three-dimensional imaging with time-based depth resolution. *Nature communications*, 7:12010, 2016.

[25] Dharmpal Takhar, Jason N Laska, Michael B Wakin, Marco F Duarte, Dror Baron, Shriram Sarvotham, Kevin F Kelly, and Richard G Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Computational Imaging IV*, volume 6065, page 606509. International Society for Optics and Photonics, 2006.

[26] DJ Townsend, PK Poon, S Wehrwein, T Osman, AV Mariano, EM Vera, MD Stenner, and ME Gehm. Static compressive tracking. *Optics express*, 20(19):21160–21172, 2012.

[27] Ashok Veeraraghavan, Ramesh Raskar, Amit Agrawal, Ankit Mohan, and Jack Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *ACM transactions on graphics (TOG)*, volume 26, page 69. ACM, 2007.

[28] Stephen S Welsh, Matthew P Edgar, Richard Bowman, Phillip Jonathan, Baoqing Sun, and Miles J Padgett. Fast full-color computational imaging with single-pixel detectors. *Optics express*, 21(20):23068–23074, 2013.

[29] Zibang Zhang, Xiao Ma, and Jingang Zhong. Single-pixel imaging by means of fourier spectrum acquisition. *Nature communications*, 6:6225, 2015.

[30] Changyin Zhou and Shree Nayar. What are good apertures for defocus deblurring? In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–8. IEEE, 2009.

[31] Changyin Zhou, Stephen Lin, and Shree K Nayar. Coded aperture pairs for depth from defocus and defocus deblurring. *International journal of computer vision*, 93(1):53–72, 2011.