# Optically multiplexed imaging with superposition space tracking

**Shikhar Uttam**[1], **Nathan A. Goodman**[1] **and Mark A. Neifeld**[1,2]
**Changsoon Kim**[3], **Renu John**[3], **Jungsang Kim**[3], **and David Brady**[3]

*Dept. of Electrical and Computer Engineering*[1], *The Optical Sciences Center*[2], *University of Arizona*

*Dept. of Electrical and Computer Engineering, Duke University*[3]

**Abstract:** We describe a novel method to track targets in a large field of view. This method simultaneously images multiple, encoded sub-fields of view onto a common focal plane. Sub-field encoding enables target tracking by creating a unique connection between target characteristics in superposition space and the target's true position in real space. This is accomplished without reconstructing a conventional image of the large field of view. Potential encoding schemes include spatial shift, rotation, and magnification. We discuss each of these encoding schemes, but the main emphasis of the paper and all examples are based on one-dimensional spatial shift encoding. System performance is evaluated in terms of two criteria: average decoding time and probability of decoding error. We study these performance criteria as a function of resolution in the encoding scheme and signal-to-noise ratio. Finally, we include simulation and experimental results demonstrating our novel tracking method.

© 2009 Optical Society of America

**OCIS codes:** (110.1758) Computational imaging; (280.0280) Remote sensing and sensors; (100.4999) Pattern recognition, target tracking

---

**References and links**

1. E. Cuevas, D. Zaldivar, and R. Rojas, "Kalman filter for vision tracking," Free University of Berlin, Tech. Rep., (2005).
2. D. J. Brady, "Micro-optics and megapixels," Optics and Photonics News **17**, 24–29, (2006).
3. D. J. Brady and M. E. Gehm, "Compressive imaging spectrometers using coded apertures," Proc. SPIE **6246**, 62460A, (2006).
4. M. A. Neifeld and P. Shankar, "Feature-specific imaging," Appl. Opt. **42**, 3379–3389, (2003).
5. D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," Proc. SPIE **6065**, 606509, (2006).
6. D. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory **52**, 1289–1306, (2006).
7. E. J. Candés, "Compressive sampling," Proc. of the Intl. Cong. of Mathematicians, (2006).
8. M. D. Stenner, P. Shankar, and M. A. Neifeld, "Wide-field feature-specific imaging," in Frontiers in Optics, Optical Society of America, (2007).
9. D. Du and F. Hwang, Combinatorial group testing and its applications, Series on Applied Mathematics **12**, (World Scientific, 2000).
10. C. M. Brown, "Multiplex imaging with random arrays," Ph.D. dissertation, Univ. of Chicago, (1972).
11. R. H. Dicke, "Scatter-hole cameras for x-rays and gamma rays," Astrophys J. **153**, L101–L106,(1968).
12. A. Biswas, P. Guha, A. Mukerjee, and K. S. Venkatesh, "Intrusion detection and tracking with pan-tilt cameras," IET Intl. Conf. VIE 06 , 565–571, (2006)
13. A. W. Senior, A. Hampapur, and M. Lu, "Image segmentation in video sequences: A probabilistic approach," WACV/MOTION'05, 433–438, (2005)

14. C. R. Wren, A. Azarbayejani, T. J. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," IEEE Trans. Pattern Anal. Mach. Intell **19**, 780–785, (1997)
15. N. Friedman and S. J. Russell, "Image segmentation in video sequences: A probabilistic approach," Proc. Uncertainty Artif. Intell. Conf., 175–181, (1997)
16. C. Stauffer and E. L. Grimson "Learning patterns of activity using real-time tracking," IEEE Trans. Pattern Anal. Mach. Intell **22**, 747–757, (2000)
17. A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 302–309, (2004).
18. K. P. Karmann and A. Brandt, "Moving Object Recognition Using an Adaptive Background Memory," in Time varying image processing and moving object recognition Volume 2, V. Cappellini (ed.), (Elsevier, 1990).
19. P. Jodoin, M. Mignotte, and J. Konrad, "Statistical background subtraction using spatial cues," IEEE Circuits Syst. Video Technol **17**, 1758–1763, (2007).
20. R. Singh, "Advanced correlation filters for multi-class synthetic aperture radar detection and classification," Carnegie Mellon University, MS Rep., (2002).
21. M. Alkanhal and B. V. K. Vijaya Kumar, "Polynomial Distance Classifier Correlation Filter for Pattern Recognition," Appl. Opt. **42**, 4688–4708, (2003).
22. B. V. K. Vijaya Kumar, "Minimum-variance synthetic discriminant functions," J. Opt. Soc. Am. A **3**, 1579–1584, (1986).
23. B. V. K. Vijaya Kumar, D. W. Carlson and A. Mahalanobis, "Optimal trade-off synthetic discriminant function filters for arbitrary devices," Optics Letters **19**, 1556–1558, (1994).
24. S. M. Kay, *Fundamentals of Statistical Processing, Volume I: Estimation Theory*. (Prentice Hall, 1993).

## 1. Introduction

There are numerous applications that require visible and/or infrared surveillance over a large field of view (FOV). The requirement for large FOV frequently arises in the context of security and/or situational awareness applications in both military and commercial domains. A common challenge associated with large FOV concerns the high cost of the required imagers. Imager cost can be classified into two categories: sensor costs and lens/optics costs. Sensor costs such as focal plane array (FPA) yield (i.e., related to pixel count), electrical power dissipation, transmission bandwidth requirements (e.g., for remote wireless applications), etc. all increase with FOV. Some of these scalings are quite severe, with process yield for example increasing exponentially with FOV. Optics costs such as size (i.e., total track), complexity (e.g., number of elements and/or aspheres), and mass also increase nonlinearly with FOV; however, these costs are somewhat more difficult to quantify without undertaking detailed optical designs. However, to illustrate the point consider two commercial lenses from Canon: the Canon EF 14mm f/2.8L II USM and the Cannon EF 50mm f/1.8 II lenses. The former is a wide-angle lens (FOV = 114 degrees) while the latter is a standard-angle lens. The wide-angle lens requires more optical elements and a sophisticated design to maintain resolution over the field of view. As a result, the wide-angle lens uses 14 optical elements and weights 645 grams while the standard-angle lens uses 6 optical elements and weights 130 grams.

We note that the various costs associated with a conventional approach to wide-FOV imaging often prohibit deployment of such imagers on platforms of interest. For example, the high mass cost together with the electrical power and bandwidth costs of conventional widefield imagers, restrict their application on many UAV platforms. Therefore, the motivation of the work reported here is to reduce the various costs of widefield surveillance, thus making it feasible for more widespread deployment.

One typical solution to this problem involves the use of narrow FOV pan/tilt cameras whose mechanical components often come with the associated costs of increased size, weight, and power consumption. The pan/tilt solution also sacrifices continuous coverage in exchange for reduced optical complexity. In most traditional approaches the goal in such problems is to reconstruct the scene of interest.

Many imaging applications, however, do not require the production of a traditional image representation (i.e., a pretty picture) suitable for human consumption. In these so-called task-

specific imagers, it is the effectiveness of the exploitation algorithm operating on the sensor output that determines overall system performance. For example, an access control imager (e.g., fingerprint sensor) may never provide image data for visual inspection. The success of this imager is entirely determined by how well it facilitates reliable matching (e.g., fingerprint recognition). Another example is tracking, in which measured image data is used to identify target locations. The success of this imager is determined only by the accuracy of the target positions that emerge from the associated postprocessing algorithm (e.g., Kalman filter [1]).

Task-specific imagers can often be compressive, in that they require many fewer measurements than the native dimension of the object space. Compressive imaging has emerged as a promising paradigm for improving imager performance and/or reducing imager resources [2], [3], [4], [5]. As a special case of compressive sensing, compressive imaging benefits from many important recent theoretical results from that domain [6], [7]. The goal of this paper is to apply concepts from compressive sensing and task-specific imaging to the problem of continuously tracking targets over a large FOV. Toward this end we propose a class of task-specific multiplexed imagers to collect encoded data in a lower-dimensional measurement space we call superposition space and develop a decoding algorithm that tracks targets directly in this superposition space. We discuss the multiplexed imagers in the next section. For now, we assume that we have this ability and briefly explain the basic idea behind superposition space tracking, which is the main focus of this paper.

We begin by treating the large FOV as a set of small sub-FOVs (disjoint sub-regions of the large FOV). All sub-FOVs are simultaneously imaged onto a common focal plane array (FPA) using a multiplexed imager. The multiple lens system of Fig. 1 is a schematic depiction of this operation. Lens shutters can also be used to control whether individual sub-FOVs are *turned on*, though for clarity the shutters are not drawn in Fig. 1. The measurement therefore is a superposition of certain sub-FOVs. A key feature of our work is applying a unique encoding for each sub-FOV, which facilitates target tracking directly on the measured superimposed data. Potential encoding schemes include (a) spatial shift, (b) rotation, (c) magnification, and/or combinations of these. These encoding methods are schematically depicted in Fig. 2.
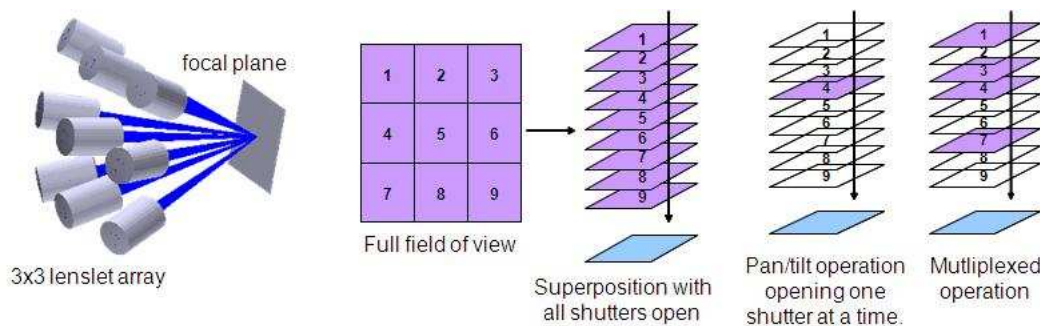


Fig. 1. Multiple lenses camera capable of performing both pan/tilt and multiplexed operations.

Encoding via spatial shifts is perhaps the easiest encoding scheme to visualize; therefore, we use this scheme for the demonstrations and performance results presented in this paper. In spatial-shift encoding, each sub-FOV is assigned a shift such that it overlaps adjacent sub-FOVs by a specified, unique amount. These spatial shifts can be one dimensional (1-D) or two dimensional (2-D). In the 1-D case, the large FOV is sub-divided along one dimension into smaller sub-FOVs. In the 2-D case, as illustrated in Fig 2(a), the full FOV is sub-divided in two orthogonal directions. Therefore, the 2-D case can be thought of as two separable 1-D cases

with decoding information shared between the two.

Rotational encoding (Fig 2(b)) assigns different rotations to each sub-FOV such that a target undergoes an angular shift in the superimposed image when it crosses between two sub-FOVs. The rotational difference between any two adjacent sub-FOVs must be unique. In a similar manner, as shown in Fig 2(c), magnification encoding assigns a unique magnification to each sub-FOV such that changes in the target's apparent size can be used to properly determine its location.

In this work we focus on 1-D spatial shift encoding due to its relative ease of implementation, easier visualization and proof-of-concept explanation, and its straightforward extension to the 2-D case. The decoding process refers to the algorithm applied to determine a target's true location in object space. We begin in Section 2 by proposing candidate optical architectures for multiplexed imagers. In Section 3 we outline the methodology for 1-D spatial shift encoding and provide a brief discussion of other encoding techniques. Section 4 provides proof-of-concept examples and explanations for the target decoding procedure while Section 5 contains demonstration and performance analysis via simulated and experimental data. We make our conclusions in Section 6.
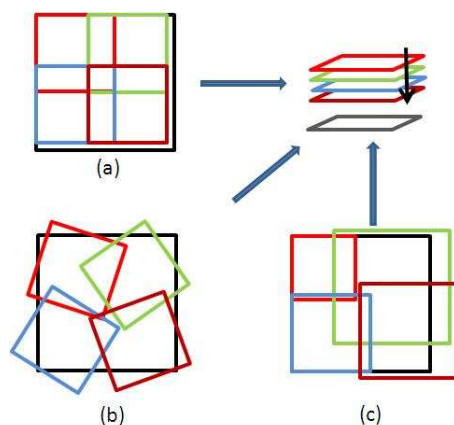
Fig. 2. Schematic representation of encoding schemes: (a) Spatial shift (two-dimensional case shown), (b) rotation, and (c) magnification.

## 2. Candidate multiplexed optical architectures

Previously we have reported on the capabilities of a novel multiplexed camera that employs multiple lenses imaging onto a common focal plane [8]. A schematic depiction of this multi-aperture camera is shown in Fig. 1. Note that each lens can have a dedicated shutter (not shown). In this camera each lens images a separate region (i.e., a sub-FOV) within the full FOV. By appropriate choice of shutter configurations, various modes of operation are possible. In one such mode all shutters are opened in sequence (one at a time), enabling an emulation of pan/tilt operation. Another mode of operation allows multiple shutters to be open at a time. This mode employs group testing concepts in order to measure various superpositions of sub-FOVs and invert the resulting multiplexed measurements to obtain a high-quality reconstructed image [2], [9], [10], [3], [11]. The third mode allows all the shutters to be open at the same time resulting in superposition of all sub-FOVs onto the common FPA.

Another optical implementation is shown in Fig. 3, in which we employ beamsplitter and mirror combinations to form the superposition measurements. This configuration reduces the lens count and avoids the image-plane tilt associated with the configuration shown in Fig. 1.

Figure 3(a) shows a setup for two sub-FOVs consisting of a beamsplitter and a movable mirror, which serves as a building block for a larger system. The optical field from the left sub-FOV ($fov_1$) is reflected by the mirror followed by the beamsplitter, and then is merged with the optical field from the right sub-FOV ($fov_2$) that has passed through the beamsplitter. The rotation of the mirror around the z-axis results in the translation of $fov_1$ along the x-direction in superposition space, providing a means to control the overlap between $fov_1$ and $fov_2$. Figure 3(b) shows an assembly of such building blocks, which can superimpose eight sub-FOVs. This implementation will serve as the basis for the experimental results presented later in sub-section 5.4.

A third implementation shown in Fig. 4 further refines the idea proposed in the second implementation by using a binary combiner in a logarithmic sequence arrangement. If we consider the same eight sub-FOVs as in Fig. 3(b), then this design allows us to access all eight sub-FOVs, using three stages of single-plate beamsplitter/mirror pairs at each stage, and three shutters placed on the mirrors. The shutters can be opened and closed in a binary sequence from 000 (all closed) to 111 (all open, superposition operation) to multiplex the eight sub-FOVs onto the camera. Although the effective aperture of the plate beamsplitter and mirror combination increases at each stage, there is an overall reduction in complexity in comparison to the optical implementation shown in Fig. 3(b). For a general scenario, when the large angular FOV is $\phi$ radians and each angular sub-FOV is given by $\beta$ radians, the number of stages in the binary combiner is given by $S = \lceil \log_2 \phi/\beta \rceil$ and the front end effective aperture $A_e$ required to avoid vignetting is approximately given by

$$A_e = 2A(1 - \tan\frac{\phi}{4})^{-1}(1-\beta)^{-S}, \qquad (1)$$

where $A$ is the aperture of the camera. To obtain this relation between effective front end aperture and the large angular FOV we fix the plate beamsplitter to be at an angle of $\pi/4$ with respect to the vertical axis and adjust the mirror to the desired angle depending on the value of $\phi$. If we define the angle of the mirror from the vertical to be $\gamma$, then for a given $\phi$, $\gamma = (\pi - \phi)/4$. This system gives us the ability to employ a camera whose angular FOV is smaller than the large angular FOV by a factor of $2^S$. We have discussed the optical scheme in a 1-D setting, but because the horizontal and vertical directions are separable, extension to 2-D is straightforward.

Figure 4 illustrates this design concept with a specific example. The angles are shown in degrees for this example. The implementation is designed for eight sub-FOVs, each having an angular range of $\beta = 7.5°$. For simplicity, the sub-FOVs are assumed to be non-overlapping, resulting in an angular FOV ($\phi$) of $60°$. The first stage folds $0°$ to $-30°$ angular range onto the $0°$ to $30°$ angular range. In the second stage the resulting $30°$ angular range is further halved to a $15°$ range using a plate beamsplitter and a mirror, which are at angles of $45°$ and $52.5°$ from the vertical axis. The third stage again halves the $15°$ range to $7.5°$ which is the range of a single sub-FOV. For the third stage the plate beamsplitter and mirror angles are $45°$ and $48.75°$, respectively. If the sub-FOVs are overlapped, then the angles of the mirrors in each stage can be adjusted to implement the given overlap.

As the angular FOV at the end of this three-stage binary combiner is reduced by a factor of $2^3$ to $7.5°$, we can use a simple lens at the end of this optical setup. We consider the TECHSPEC MgF2 coated achromatic doublet with a diameter of 12.5mm and a focal length of 35mm. Given $A = 12.5$mm, using (1) we calculate the front end effective aperture $A_e$ of the beamsplitter and mirror combination to be 5.2cm. Since the optical system does 2-D imaging we have the same three-stage binary combiner in the other dimension with the same effective front end aperture. In total therefore, we have six plate beamsplitter and mirror combinations. Assuming, for simplicity, that the beamsplitter and the mirror equally share the effective aperture, the lengths of the plate beamsplitter and the mirror are given by $(5.2/2)\sqrt{2}$cm and $(5.2/2)(2/\sqrt{3})$cm. The factors of $\sqrt{2}$ and $2/\sqrt{3}$ follow from the plate beamsplitter and the mirror being at angles of

45° and 30° respectively from the vertical axis . Assuming a square size for both, the Stage 3 dimensions of the plate beamsplitter are 3.7cm × 3.7cm and that of the mirror are 3cm × 3cm. Also assuming the thickness of the optical glass to be 2mm, we calculate that in Stage 1 the total volume of glass used by the two pairs (corresponding to 2-D imaging) of plate beamsplitter and mirror combination is 9.1cm$^3$. Doing similar calculations for Stages 2 and 3 gives the volume of glass used to be 6.8cm$^3$ and 4cm$^3$, respectively. If we take the density of the optical glass to be 2.5gm/cm$^3$, the total mass of the log combiner is 49.75gm. The mass of the achromatic doublet is less than 5gm and hence the weight of the optical system is about 54.75gm. If we were to directly use a wide-angle lens to capture an angular FOV of 60°, then a potential candidate is Canon's EF 35mm f/1.4L USM lens. It has an angular FOV of 63°, but its mass is 580 gm and it uses 11 optical elements. Therefore, we see savings of about a factor of 10 for our proposed multiplexed imager and also reduced optical complexity as we are using a simple, easy-to-design binary combiner and an achromatic doublet as opposed to an 11-element wide-angle lens.
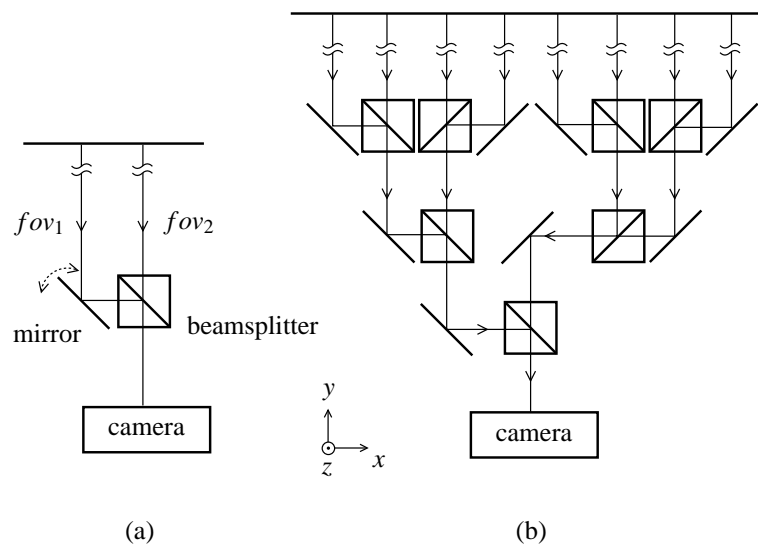


Fig. 3. Optical setup capable of performing multiplexed operations with spatial shift encoding for (a) $N_s = 2$ and (b) $N_s = 8$

Two practical issues that arise in designing optical systems involving beamsplitters are vignetting and transmission loss. Vignetting arises when there is non-uniformity in the amount of light that passes through the optical system for each of the points in the FOV. This resulting non-uniformity at the periphery of the superposition data has the potential of increasing false negatives which in turn can lead to errors in properly locating the targets. To overcome this potential problem in the setup shown in Fig. 3, the size of each beamsplitter should be large enough to ensure that the angular range subtended by that beamsplitter at the camera is a strict upper bound on the angular range of the corresponding sub-FOV. Field stops are then used to restrict the angular range of the beamsplitter to that of the sub-FOV. Specifically, for the setup shown in Fig. 3 and used in Section 5.4, each sub-FOV is 1.9° horizontally and 1.3° vertically while the angular range of the beamsplitter is approximately 3°. As a result, vignetting is avoided. For the binary combiner shown in Fig. 4 vignetting is not an issue because (1) is derived from a vignetting analysis of the binary combiner, to give an effective aperture $A_e$ that does not block light from any point in the large angular FOV.
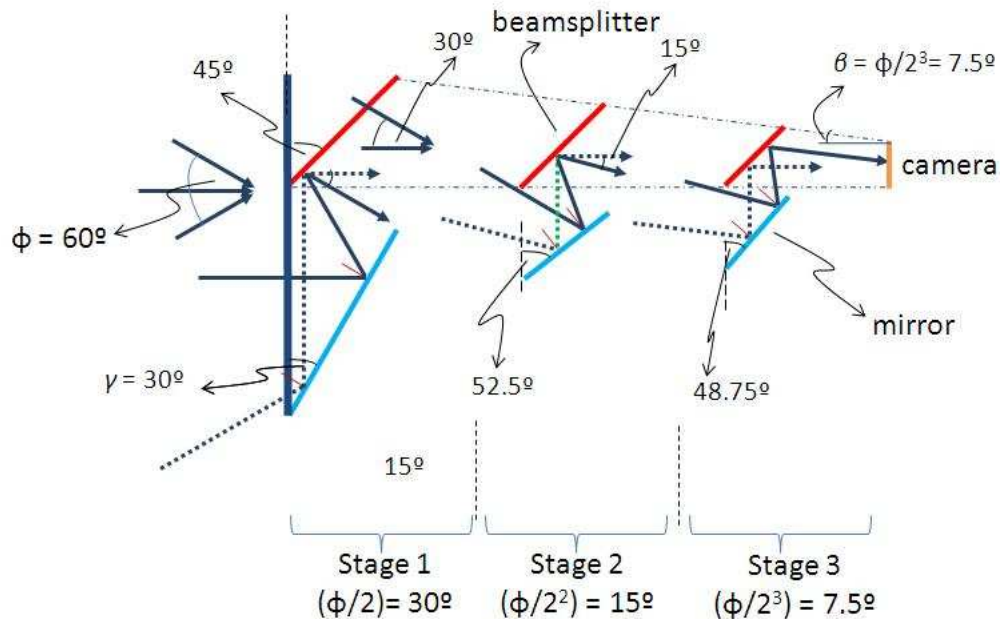
Fig. 4. Binary combiner in a log sequence arrangement for multiplexing 8 sub-FOVs each with an angular range of 7.5°.

The second issue has to do with a decrease in throughput due to optical "combination loss" when the light passes through the various stages of the beamsplitters. Specifically for the eight sub-FOV multiplexed imager shown in Figure 3(b) the optical transmission decreases by $(0.5)(0.5)(0.5) = 0.125$. This throughput cost, however, is no worse than that of a narrow-field imager in a pan-tilt mode which is commonly used to achieve a wide FOV [12], [13]. For our current example the dwell time of a corresponding narrow-field imager on each sub-FOV will be $1/N_s = 1/8 = 0.125$ time units. Since the photon count is directly proportional to the dwell time, we have the same photon efficiency for both the multiplexed and wide-field imagers. This result can be extended to a general case of $N_s$ sub-FOVs where the photon efficiency of both the multiplexed and wide-field imagers is reduced by a factor of $N_s$. We acknowledge, however, that for the proposed beam-splitter and mirror-based multiplexed imagers we have not taken the component losses, e.g. scattering at the mirror, into account. We assume these losses to be small in comparison to the photon efficiency. Unlike the beamsplitter and mirror based multiplexed imagers, the multiple-lens-based imager shown in Fig. 1 overcomes the disadvantage of loss in photon efficiency at the expense of a higher optical mass cost.

The imagers we have proposed are used to simultaneuosly encode and compress (through superposition of the sub-FOVs) the data. The subsequent algorithm then performs target tracking by decoding the relevant information from this compressed superpositon data. In the next section we discuss in detail the need for encoding and also explain the encoding methodology with respect to 1-D spatial shifts.

## 3. Sub-FOV superposition and encoding

As discussed in the Introduction, our goal is to track targets over a large FOV. We suppose this large FOV to be $H$ distance units in the vertical dimension by $W$ distance units in the horizontal (encoded) dimension. The 1-D spatial shift encoding strategy is best understood by considering

the following three domains or spaces we work with: object space, superposition space, and hypothesis space.

The object space is the full area on the ground that is actually observed by the sensor. For sake of clarity, we begin by letting the sub-FOVs be uncoded. Uncoded sub-FOVs are obtained when the $H \times W$ area of the large FOV is simply divided into $N_s$ adjacent sub-FOVs, without overlap, as shown in Fig. 5(a). Each of the resulting sub-FOVs is $H \times W_{fov}$ in size where $W_{fov} = W/N_s$. Using an optically multiplexed imager, we image each of these sub-FOVs onto a common FPA. If the object space resolution of the optical system is $\Delta r$ distance units in each direction, then the dimensionality (in pixels) of a single sub-FOV will be $H/\Delta r$ pixels by $W_{fov}/\Delta r$ pixels.

Optical multiplexing of all sub-FOVs onto an FPA the size of a single sub-FOV corresponds to superposition of all the sub-FOVs. Thus, the measured image comprises what we call the superposition space. Note that the superposition of all sub-FOVs onto a single FPA provides data compression - $H/\Delta r$ by $W/\Delta r$ pixels are imaged using an $H/\Delta r$ by $W_{fov}/\Delta r$-pixel FPA. Specifically, the compression ratio for the uncoded case being discussed is $N_s$. Measuring only the superposition, however, introduces *ambiguity* into the target tracking process. Consider a single target moving through the first sub-FOV in the uncoded object space as shown in Fig. 5(a). The corresponding superposition space looks like Fig. 5(b). Based on the encoding used (in this case none) and the size of a single sub-FOV we decode possible locations of the target in object space. We call this new space the hypothesis space (see Fig. 5(c).) The hypothesis space is *not* a reconstruction of the object space - it is a visualization of the decoding logic.

The single target detected in the superposition space of Fig. 5(b) does not provide information about the true sub-FOV where the target is located. Therefore, there is ambiguity in the hypothesis space, and we hypothesize that the target could be in any of the $N_s$ sub-FOVs. In fact, for this uncoded case, it is not possible to correctly decode the target location based on measurements in superposition space. To overcome this ambiguity, we need a distinguishing characteristic that appears in the superposition space yet identifies a target's true location in object space. This trait can be provided by encoding the sub-FOVs with a spatial shift in the object space. Instead of defining non-overlapping sub-FOVs, we allow for some overlap between adjacent sub-FOVs in the object space as seen in Fig. 6(a). In this shift-encoded system, when a target passes through an area of overlap between two or more sub-FOVs, instead of a single target being present in superposition space there are multiple copies of the original target as shown in Fig. 6(b). We refer to these multiple copies as *ghosts* of the target. The presence of these ghosts allows the target location to be decoded as long as the pairwise overlap between adjacent sub-FOVs is unique.

The overlaps can be designed in different ways. They can be random and unique such that no two overlaps are the same, or they can be integer multiples of a fundamental shift. Also, these integer multiples need not be successive, but can be randomly picked from the available set. The simplest design, however, is to let the overlaps be successive multiples of a fundamental shift, which we call the shift resolution $\delta$. Given that there are $N_s$ sub-FOVs, we define the unique overlaps to be the overlap set $\mathbf{O} = \{0, \delta, 2\delta, \ldots, (N_s - 2)\delta\}$. We can now construct a 1-D spatial-shift encoded object space as follows:

1. Start with the uncoded sub-FOVs.

2. Let the first two (from the left) sub-FOVs be in the same position (non-overlapped) as in the uncoded case. We label these sub-FOVs as $fov_0$ and $fov_1$ respectively;

3. Shift the $i^{\text{th}}$ sub-FOV ($fov_i$), $i = \{2, \ldots, N_s - 1\}$ to the left such that it overlaps with the $(i-1)^{\text{th}}$ sub-FOV by $O_{i-1}$. Depending on the size of the overlap, it is possible that the shift causes portions of $fov_i$ to overlap with more than one sub-FOV. (Figure 7 is an example.) One condition that must be satisfied is that $fov_i$ cannot completely overlap
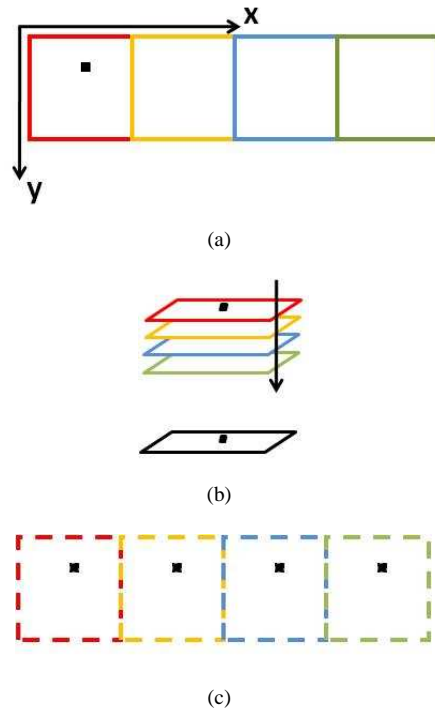
Fig. 5. (a) The area of interest (large FOV) for tracking targets along with the reference coordinate system. The large FOV is sub-divided into 4 non-overlapping sub-FOVs in the x-direction. In this un-coded case the object space and the FOV are the same. (b) Super-position space. (c) Hypothesis space: From the superposition space it is not possible to tell which sub-FOV the target belongs to. This ambiguity in target location leads to the hypothesis that the target could be in any of the 4 sub-FOVs.

$fov_{i-1}$. This condition implies a restriction on the shift resolution $\delta$. The shift resolution can range from zero (uncoded) to a maximum of $\delta_{max} = W_{fov}/(N_s - 1)$.

As shown in Section 4, the resulting encoded object space enables the target location to be decoded. The disadvantage is that the object space now covers a smaller area than the uncoded case. The object space is largest when $\delta = 0$, which corresponds to an uncoded case with target decoding ambiguity. The object space is smallest (covers the least area) when $\delta = \delta_{max}$. Between these two extremes, there is a compromise between area coverage and the smallest shift resolution that must be detected in the decoding procedure.

To quantify the reduction in area coverage, we define area coverage efficiency ($\eta$) to be the ratio between the area of the encoded object space and the uncoded object space. The shift resolution also affects the compression ratio ($r$), which is defined as the ratio of the area of the encoded object space to the area of a single sub-FOV. The area coverage efficiency and the compression ratio are given by

$$\eta = 1 - \alpha \frac{(N_s - 2)}{2N_s}, \tag{2}$$

$$r = N_s - \alpha \frac{(N_s - 2)}{2}, \tag{3}$$

where $\alpha = \delta/\delta_{max}$ lies between 0 and 1. Small $\alpha$ results in better area coverage and higher compression ratio but smaller shift resolution. In addition, if we define a boundary as a line in object space where there is a change in the sub-FOV overlap structure, then small alpha also results in a lower boundary density, which can adversely affect the average time required to properly decode a target. The opposite characteristics are true for values of $\alpha$ near unity. We quantify trade-offs between area coverage, decoding accuracy, and average decoding time in more detail in Section 5.

### 3.1. 2-D spatial, rotational and magnification encodings

Although 1-D spatial shift encoding is the main focus of this paper, we now take time to briefly describe other potential encoding schemes.

As mentioned previously, 2-D spatial shift encoding can be thought of as two separable 1-D spatial shift encodings. Specifically, instead of sub-dividing the large FOV into smaller sub-FOVs in only the $x$-direction, we also sub-divide in the $y$-direction. Again starting with the uncoded case, if the number of sub-FOVs in the two directions are $N_x$ and $N_y$ respectively, the size of each resulting sub-FOV is $H_{fov} \times W_{fov}$ where $H_{fov} = H/N_y$ and $W_{fov} = W/N_x$. Defining the shift resolutions in the two dimensions as $\delta_x$ and $\delta_y$, the overlap sets are then given by $\mathbf{O}_x = \{0, \delta_x, 2\delta_x, \ldots, (N_x-2)\delta_x\}$ and $\mathbf{O}_y = \{0, \delta_y, 2\delta_y, \ldots, (N_y-2)\delta_y\}$. The encoding procedure described above for a 1-D system can be separably applied to the sub-FOVs in both the $x$- and $y$-directions to give the 2-D spatially encoded object space. In this 2-D encoding scheme, each sub-FOV is characterized by a unique pairwise combination of horizontal overlap from $\mathbf{O}_x$ and vertical overlap from $\mathbf{O}_y$. The area coverage efficiency and the compression ratio are given by

$$\eta = (1 - \alpha_x \frac{(N_x - 2)}{2N_x})(1 - \alpha_y \frac{(N_y - 2)}{2N_y}), \tag{4}$$

$$r = (N_x - \alpha_x \frac{(N_x - 2)}{2})(N_y - \alpha_y \frac{(N_y - 2)}{2}), \tag{5}$$

where $\alpha_x = \delta_x/(\delta_x)_{max}$, $\alpha_y = \delta_y/(\delta_y)_{max}$, and $(\delta_x)_{max}$ and $(\delta_y)_{max}$ are the maximum allowable shifts in the two dimensions.

In the case of rotational encoding the objective is to define unique rotational differences between any two sub-FOVs. The simplest way to do this is to define a fundamental angular resolution ($\delta_{ang}$) and let all the rotational differences be integer multiples of ($\delta_{ang}$). The resulting rotational difference set is $\mathbf{R} = \{0, \delta_{ang}, 2\delta_{ang}, \ldots, (N_s-1)\delta_{ang}\}$. One must be careful, however, to note that $\mathbf{R}$ is a set of rotational *differences*, i.e., the difference between the absolute rotations of any two sub-FOVs. The absolute rotation assigned to the $i^{\text{th}}$ sub-FOV is then $\sum_{j=0}^{i} R_j, i = 0, 1, \ldots, N_s - 1$. Furthermore, since rotational encoding is periodic every 360°, it is logical to upper bound the maximum absolute rotation by 360°. This bound results in a maximum angular resolution of $\delta_{max} = 2 \cdot 360°/(N_s(N_s - 1))$. Rotational encoding like spatial shift encoding can be applied to sub-FOVs in either or both $x$- and $y$- directions. We call rotational encoding 1-D when the large FOV is sub-divided in either $x$- or $y$- direction and rotational encoding is then applied to the resulting sub-FOVs. 2-D rotational encoding refers to the case where rotational encoding is applied to the sub-FOVs resulting from the sub-division of the large FOV in both the directions. Unlike 2-D spatial shift encoding, however, 2-D rotational encoding is not separable. It requires that the rotational difference between any two sub-FOVs, regardless of the dimensions they lie along, be unique. In 2-D spatial shift encoding on the other hand, overlaps have to be unique only with respect to one direction. As a result even when $\mathbf{O}_x = \mathbf{O}_y = \mathbf{O}$ and $N_x = N_y = N_s$, the resulting 2-D spatial shift encoding is valid because each sub-FOV will still have a unique overlap $(O_i, O_j), i, j \in 0, 1, \cdots, N_s - 1$ associated with it.

Magnification encoding assigns unique magnification factors to each sub-FOV. The magnification factors depend on the optical architecture and the size of the area of interest. 2-D magnification encoding, like 2-D spatial shift encoding, is separable as long as we can separably control the magnification factors along the two directions. We can now define sets of unique magnification factors $\mathbf{M}_x$ and $\mathbf{M}_y$ along the $x$- and $y$-directions respectively. This results in an encoding scheme similar to the 2-D spatial shift encoding. As a result, in a manner analogous to 2-D spatial shift encoding, even when $\mathbf{M}_x = \mathbf{M}_y = \mathbf{M}$ and $N_x = N_y = N_s$, the magnification encoding scheme is valid because each sub-FOV will still have a unique combination $(M_i, M_j), i, j \in 0, 1, \cdots, N_s - 1$ of horizontal and vertical magnification factors applied to it. Finally, we note that it may be possible to obtain greater area coverage for the same FPA by combining several encoding methods.
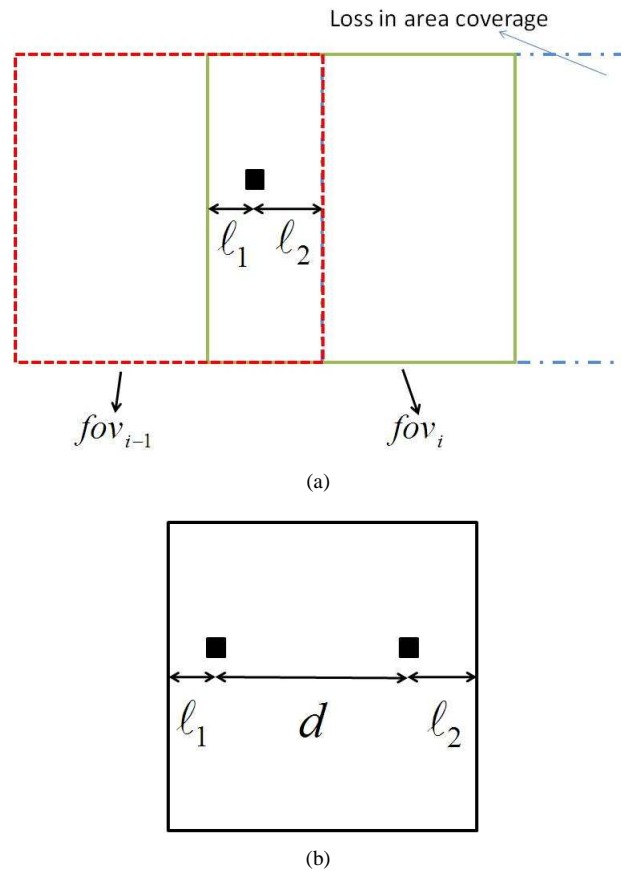


Fig. 6. (a) A portion of an encoded object space with the target in the region of overlap between the two adjacent sub-FOVs along with its distance from the two boundaries of the overlap. The loss in area coverage due to encoding is also shown. (b) Superposition space with two target copies or ghosts. The one to the left corresponds to $fov_i$ and the one to the right corresponds to $fov_{i-1}$. Also depicted is the relation between the distance measures $\ell_1$ and $\ell_2$ in the object space, and the separation distance between the target ghosts in the superposition space.

## 4. Decoding: proof of concept

Properly encoded spatial shifts enable decoding of the true target location whenever the target crosses a boundary into a new overlap region, and possibly sooner. Depending on the sub-FOV shift structure, target replicas or ghosts can appear only at certain fixed locations, that is, the distance in superposition space between any set of ghosts corresponding to the same target will have a unique pattern because each sub-FOV overlap is unique. From this unique pattern, we can identify the sub-FOVs involved and uniquely localize the target in hypothesis space. Once the target is uniquely located in hypothesis space, we have decoded the target's position in object space.

In sub-sections 4.2 and 4.3 we explain and demonstrate the decoding procedure - first for a single target and then for multiple targets. We begin though with a brief discussion on correlation based tracking employed in this work.

### 4.1. Correlation-based tracking

In the simulations and performance analyses we use a Kalman filter to track target ghosts in the superposition space. Our Kalman tracker has a length-four state vector, the four states being the $x$- and $y$-coordinates, and the $x$- and $y$-direction velocities of detected target ghosts. Kalman tracking involves two basic steps: (1) updating the estimate (mean and covariance) of the state vector at time $t$ based on new measurements made at time $t$, and (2) propagating forward the revised estimate at time $t$ to time $t + 1$. We use correlation to make these new measurements for the update step. Correlation performs three specific tasks: (1) locating the target ghost positions - and as a result, their target velocities - in the superposition space (measured data), (2) separating them into different classes in case of multiple targets (multiple tracks), and (3) separating weak target ghosts from noise and background clutter. If the target ghosts have a strong signal-to-noise ratio (SNR), then the first two steps can be performed using template matching. On the other hand, if the signal strength is weak, template matching will not suffice. This is an important point since in our proposed technique superposition of the sub-FOVs results in a reduction of the target ghost's dynamic range. Specifically let us consider $N_s$ sub-FOVs, each with a dynamic range of $[0, R]$. Let the target be present in only one sub-FOV and in a region that does not overlap with adjacent sub-FOVs. When we superimpose these sub-FOVs, the dynamic range of the resulting superposition space, neglecting saturation, is $[0, N_s R]$ while the dynamic range of the target is still $[0, R]$. Therefore, the target strength is suppressed by a factor of $N_s$. If the target is in a region of overlap of $M$ sub-FOVs, $M < N_s$, then this factor is reduced to $N_s/M$.

The above analysis shows that there is a trade-off between the dynamic range of the target, the number of sub-FOVs ($N_s$) and the size of the object space. (For a given $N_s$, the object space size is related to the number of overlaps $M$ and the shift resolution $\delta$, and therefore, affects the dynamic range.) This necessitates that our system be able to deal with presence of weak targets. There is a two-fold strategy we can consider. First, use a statistical background subtraction technique to remove the background. Second, use correlation filters to locate and classify target ghosts. We briefly look at each of these methods.

Background subtraction is an intuitive way to reduce background clutter and thereby increase target strength. This directly helps us in increasing the dynamic range of the targets. In almost all real life cases, the background is non-static. To faithfully estimate the non-static background we use statistical subtraction techniques. Depending on the complexity of the background, the background image pixels are modelled as having Gaussian probability density functions (pdfs) [14] or as mixture of Gaussians (MoG) [15], [16]. If the parametric Gaussian mixtures do not embody enough complexity then kernel based methods can also be used [17]. All these methods fall under the category of non-predictive methods. Predictive methods employ-

ing Kalman tracking based techniques to characterize the state of each pixel are also used for background estimation [18]. The biggest challenge with both the predictive and non-predictive methods is that they require a time sequence of image frames having no target motion to characterize the background. This, however, is not possible in many real-time scenarios. Recently Jodoin et al. [19] proposed a novel spatial approach to background subtraction which works under the assumption of ergodicity: temporal distribution observed over a pixel corresponds to the statistical distribution observed spatially around that same pixel. They model a pixel using unimodal and multimodal pdfs and train this model over a single frame. This method allows us to estimate the background from a single image frame which results in a faster algorithm that requires less memory.

Background subtraction removes background clutter, but not necessarily the noise present in the measured data. To further increase robustness against this residual noise we employ advanced correlation filters for making state vector measurements in the update step of our Kalman tracker. Correlation filters, because of their shift invariance and their distortion tolerance ability, have been successfully employed in radar signal processing and image analysis for pattern recognition [20], [21]. Correlation filters such as minimum-variance synthetic discriminant functions (MVSDF) [22] and optimal trade-off synthetic discriminant function (OTSDF) filter [23] show good ability for detecting and classifying multiple targets in the presence of noise and background clutter.

Although we do not focus on one particular system design in this paper, we acknowledge that in the presence of noise and background clutter, the ability to make accurate measurements is an important step in our proposed target tracking approach, and the above mentioned techniques of statistical background subtraction followed by advanced correlation filters for detecting and locating targets provide us this ability.

### 4.2. Decoding procedure for a single target

Consider a target in object space that enters the region of overlap $O_i$ between two sub-FOVs, $fov_{i-1}$ and $fov_i$, as shown in Fig. 6(a). The corresponding superposition space looks like Fig. 6(b). Under the assumption of a single target, the presence of two ghosts in the superposition space indicates that the target has entered a region where two sub-FOVs overlap. To find the two sub-FOVs creating this overlap, we first measure the distance between the two ghosts in superposition space. Let this distance be $d$. Based on knowledge of the shift encoding, we then calculate the set of all possible separations between two ghosts of a single target. We label this set $\mathbf{S}$ and call it the separation set. Recalling that the set of adjacent sub-FOV overlaps is the overlap set $\mathbf{O}$, the elements of the separation set are $S_i = W_{fov} - O_i, i = 0, 1, \ldots, N_s - 2$. The set $\mathbf{S}$ can be computed once in advance and stored for future reference.

We can now define the set $\mathbf{T}_2 \subseteq \mathbf{S}$ such that it contains only those elements of $\mathbf{S}$ which are realizable in the spatial shift encoding scheme. It is important to note that not all elements of $\mathbf{S}$ result in a valid element of $\mathbf{T}_2$. It is possible that the region of overlap between two sub-FOVs lays within (i.e., is a sub-region of) the region of overlap between the two mentioned sub-FOVs and one or more of other sub-FOVs. In such a case a target present in the two sub-FOV overlap region will always produce more than two ghosts in the superposition space. An example of this scenario is given in Section 5. The subscript '2' in $\mathbf{T}_2$ indicates that the target is in the region of overlap between two sub-FOVs only (as opposed to regions covered by more than two sub-FOVs).

We now look at the basic principle for decoding a target in the region of overlap between two sub-FOVs. In Fig. 6(a), the distances $\ell_1$ and $\ell_2$ are the distances from the edges of the two overlapping sub-FOVs. In Fig. 6(b), we see that these distances are the same as the distances from the two ghosts to the edge of the superposition space. Since $\ell_1 + \ell_2$ is equal to the overlap

between the sub-FOVs, $\ell_1 + \ell_2$ is an element of **O**. If the measured separation $d$ corresponds to the $j^{\text{th}}$ element of $\mathbf{T}_2$, then we can decode $fov_{i-1}$ as the $j^{\text{th}}$ sub-FOV and $fov_i$ as the $(j+1)^{\text{th}}$ sub-FOV. Finally, our *a priori* knowledge of the sub-FOV locations in object space along with the position of the corresponding target ghosts in superposition space can now be used to decode the target's $x$-coordinate in hypothesis space. Because we have used 1-D spatial shift encoding, the $y$-coordinate of the target is the same in superposition space, hypothesis space, and object space. We have now completely decoded the target location.

The above example considers two overlapping sub-FOVs. We can extend the decoding procedure to the case where a single target enters a region where more than two sub-FOVs overlap. Such a scenario can arise, for example, when the overlaps $O_{i-1}$ and $O_i$ are such that $fov_{i+1}$ not only overlaps with $fov_i$, but also with $fov_{i-1}$ as in Fig. 7 where sub-FOVs $fov_2$, $fov_3$ and $fov_4$ overlap. In such cases the number of target ghosts appearing in superposition space will be equal to the number of overlapping sub-FOVs covering the target. In general, assuming this number to be $M$, we first calculate the sequence of pair-wise distances, from left to right, between the $M$ target ghosts in superposition space. This sequence is referred to as *target pattern* $d_M$. We then compare the sequence to the allowed length-$M$ target patterns in superposition space, which again are known because the shift encoding structure is known. The set of allowed length-$M$ target patterns is denoted by $\mathbf{T}_M$. The matching pattern determines the proper set of $M$ sub-FOVs, from which the target's position in hypothesis space can be fully decoded as in the case above for two overlapping sub-FOVs.

### 4.3. Decoding procedure for multiple targets

When our above mentioned two-fold strategy is able to associate target ghosts with the correct target for multiple targets, we can simply apply the decoding procedure for a single target to all the targets iddividually. On the other hand, when we have scenarios where the targets have (1) identical or only subtle shape differences, or (2) such a weak signal strength that only target detection is possible, we need a way to associate the target ghosts with the correct targets. In such scenarios where direct associations are not possible, we need a procedure for decoding multiple targets. The proposed procedure is essentially the same as for a single target except for a pre-decoding step where ghosts in superposition space belonging to the same target are associated with each other. The procedure involves the following indirect three-fold strategy (stated here specifically with respect to 1-D shift encoding):

1. Group all detected targets in superposition space according to their $y$-coordinate values. Since the system has 1-D spatial shift encoding in the $x$-direction, ghosts belonging to the same target must have the same $y$-position. However, it is possible for two different targets to also have the same $y$-coordinate. Therefore;

2. Compare the estimated velocities of potential targets in each group. If multiple velocities are detected, it is assumed that multiple targets are present, and the group is sub-divided. This step follows from the observation that ghosts belonging to the same target must have the same (2-D) velocity. Members of each target group now have the same velocity and the same $y$-coordinate. Finally;

3. Begin decoding process by comparing allowed target patterns to the target patterns of groups determined in the first two steps. The allowed target patterns are the sets $\mathbf{T}_i, i = 2, 3, \ldots, K$, where $K$ is the maximum number of sub-FOVs that overlap. We begin with the highest order target patterns ($\mathbf{T}_K$) and work down to the lowest order target patterns ($\mathbf{T}_2$). When a pattern is detected, the target position is decoded. If an allowed target pattern is not detected, the targets in the group are assumed to reside in regions of object space without overlapping sub-FOVs.
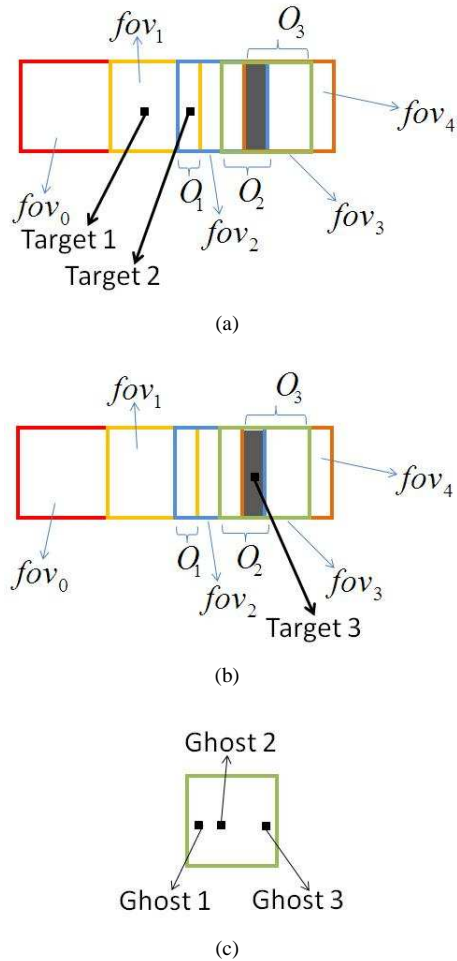
Fig. 7. Three targets in the object space with the same velocities and *y*-coordinates depicting two scenarios both of which result in the same superposition space data. (a)*Scenario 1*: Two targets in the object space with Target 1 in the non-overlapping region of $fov_1$ and Target 2 in the region of overlap between $fov_1$ and $fov_2$. (b) *Scenario 2*: A single target (Target 3) in the overlap between the sub-FOVs $fov_2$, $fov_3$ and $fov_4$. (c) The same superposition space arising from the two scenarios. (i) Ghost 1 and Ghost 3 in the superposition space are ghosts of Target 2 in the object space, while Ghost 2 in the superpostion space corresponds to Target 1 in the object space. (ii) Ghost 1 through 3 in the superposition space are ghosts of Target 3 in the object space.

When performed in order, these steps usually enable decoding of the locations of multiple targets. Under certain conditions, however, correct decoding is not possible. Figure 7 shows two scenarios, the first (Fig. 7(a)) involving two targets as was explained above and the second (Fig. 7(b)) involving a single target in a region with three overlapping sub-FOVs. On the rare occassions when this occurs and the targets involved happen to have the same *y*-coordinate and estimated velocity, it is not possible to decode which scenario is the true scenario (Fig. 7(c)) and, according to the decoding rules above, the higher order shift pattern will be decoded. (In Fig. 7, scenario #2 will be decoded). In Fig. 8 we illustrate this general case with an example

movie. Each movie frame shows the object space across the top, the superposition space in the middle, and the hypothesis space across the bottom. Object space represents the "truth" while the superposition space represents the actual measurement data. The hypothesis space visualizes how the decoding logic works in real time. We stress that the hypothesis space is *not* a reconstruction of the object space, but is simply a visualization of the decoding logic. The "truth" background has been added to the hypothesis space simply to provide visual perspective to the viewer. The movie shows two targets with the same (2-D) velocity and the same *y*-coordinates moving through the object space. By unfortunate coincidence the targets happen to have a horizontal separation which is an element of set $\mathbf{T}_2$. As a result, the two targets are decoded as a single target, and their decoded position jumps around. Eventually, the velocities of the two targets change, the superposition space ghosts are properly grouped, and the two targets are correctly decoded. We would like to remind the reader that here we are assuming the targets have either identical shape or have such subtle differences in shape that association based on shape is not possible or reliable. We will continue to make this assumption throughout the rest of the paper.
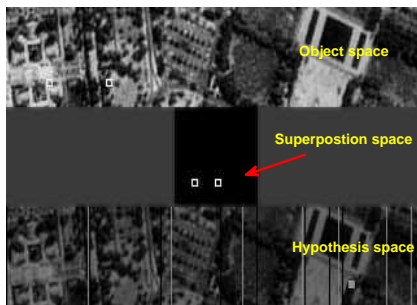


Fig. 8. (Media 1) showing error in decoding two targets with the same velocity, the same *y*-coordinate, and with a separation which is an element of set $\mathbf{S}_2$. Correct decoding is possible only when the targets begin to differ in their velocities. (Movie size: 0.9 MB)

### 4.4. *Decoding via missing ghosts*

Thus far, we have described how the difference, or shift, between target ghost positions can be used to properly decode target location by uniquely identifying the overlap region that must have produced the shift. However, additional decoding logic is available to the sensor in the form of *missing ghosts*. Although this additional logic may not be able to uniquely decode the target, it reduces the number of target locations in hypothesis space. To explain this principle, consider the following scenario.

For clarity, we focus on a single target moving through the object space. We also assume that $N_s = 4$ with the sub-FOVs encoded according to 1-D shifts belonging to the overlap set $\mathbf{O} = \{0, \delta, 2\delta\}$. The target is in $fov_0$ and is moving towards $fov_1$ in the object space as illustrated in Fig. 9(a). Since there is zero overlap between $fov_0$ and $fov_1$, superposition space has a single target as seen in Fig.9(b). Based on the superposition space measurement and the decoding strategy explained above, the target cannot be completely decoded. We can only hypothesize a potential target location in each of the four sub-FOVs. Therefore, hypothesis space looks like Fig. 9(c). The local *x*-coordinate of each hypothesized target in its respective sub-FOV is the same. However, we can apply our knowledge of overlaps to rule out $fov_2$. This sub-FOV cannot be allowed because if the target were truly at this position, it would imply that the target
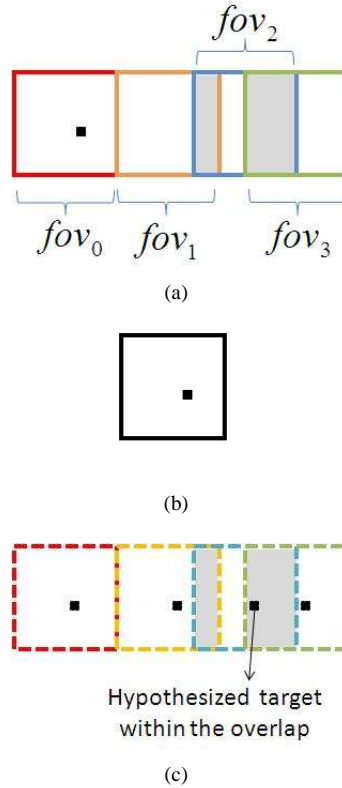
Fig. 9. (a) Encoded Object Space with 4 sub-FOVs. The target is in $fov_0$. The local $fov_0$ x-coordinate lies between $W_{fov} - O_2$ and $W_{fov}$. (b) The superposition Space with a single target indicating that the target is in a non-overlapping sub-FOV region. (c) Hypothesis Space with 4 potential targets. The third hypothesized target, however, lies in a sub-FOV overlap which if true would produce ghosts in the Superposition Space. The absence of ghosts rules out this hypothesized target as a potential true target.

resides in an overlapping region between $fov_2$ and $fov_3$. The absence of a second ghost in superposition space tells us this is not true. Therefore, the target can only be in $fov_0$, $fov_1$ or $fov_3$, and we have reduced the target hypotheses from 4 to 3. If the target continues to move toward $fov_2$, additional sub-FOVs will be ruled out by the same logic. In general, missing ghosts can be used to rule out anywhere from one to all incorrect sub-FOVs depending on the target location and encoding structure.

## 5. Results

To demonstrate and quantify the efficacy of the proposed 1-D spatial shift encoding scheme, we present results generated from both simulated data and a laboratory experiment.

### 5.1. Simulation

We simulated an object space with $N_s = 8$, 1-D spatial-shift encoded sub-FOVs. The size of each sub-FOV was $64\Delta r$ distance units by $64\Delta r$ distance units, where $\Delta r$, the object space resolution, was assumed to be finer than the size of the targets of interest. The corresponding sub-FOV dimensionality (in pixels) is 64 by 64. In the simulation we used template matching to
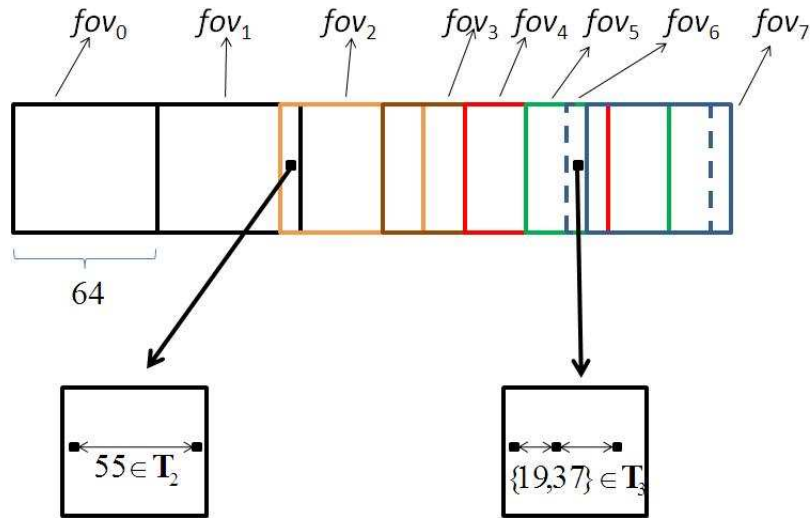
Fig. 10. Examples of valid target patterns.

perform correlation for locating target ghost positions. In order to subtract the background clutter from the measured data before performing the correlation, we characterized the background by averaging the superposition space data for a large number of frames under the assumption that we have the ability to observe the scene for a long time with respect to the frame rate.

We first defined the overlap set as $\mathbf{O} = \{0, 1\delta, 2\delta, 4\delta, 3\delta, 5\delta, 6\delta\}$ where $\delta$ was chosen as $\delta = \delta_{max} = floor(W_{fov}/(N_s - 1)) = \text{floor}(64/7) = 9$ pixels. The reason for this choice of $\delta$ was that large $\delta$ increases the boundary density in the superposition space, which increases the total overlapped area. As a result, the number of target ghosts in superposition space that need to be tracked increases. If the decoder and tracker can handle a large number of targets in superposition space, we can gain confidence that the decoding procedure is robust. Note that the overlaps in the set $\mathbf{O}$ are not monotonically increasing, which shows that the order of the overlaps in 1-D shift encoding is arbitrary. Based on this overlap set, the separation set is computed to be $\mathbf{S} = \{64, 55, 46, 28, 37, 19, 10\}$. The allowed target patterns are then given by $\mathbf{T}_2 = \{64, 55, 46, 28, 37, 10\}$, $\mathbf{T}_3 = \{\{37, 28\}, \{19, 37\}, \{10, 19\}\}$ and $\mathbf{T}_4 = \{\{10, 19, 37\}\}$. The sets of overlapping sub-FOVs corresponding to these patterns are $\mathbf{F}_2 = \{\{0, 1\}, \{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{6, 7\}\}$, $\mathbf{F}_3 = \{\{3, 4, 5\}, \{4, 5, 6\}, \{5, 6, 7\}\}$ and $\mathbf{F}_4 = \{\{4, 5, 6, 7\}\}$. It is important to point out that $\{5, 6\}$ is absent from the set $\mathbf{F}_2$. This does not imply that the two sub-FOVs do not overlap, which they do, but instead means that the region of overlap between $fov_5$ and $fov_6$ is also overlapped by a third or a even a fourth sub-FOV as shown in the sets $\mathbf{F}_3$ and $\mathbf{F}_4$ respectively. In Fig. 10 we illustrate the allowed target patterns with a couple of examples.

We simulated a scenario by populating the object space with four identically shaped targets appearing at random locations, with random velocities, at random times, and lasting for random durations. We allowed the starting target locations to be anywhere in the object space with equal probability. The velocities were uniformly distributed between 0 and $3\Delta r$ distance units for every time step to ensure that the target movement looked real. The start and stop times were uniformly distributed between 0 and 100 time steps. Figure 11 shows a movie of one such example simulation which best illustrates all the facets of our decoding procedure. We explain this movie in the next paragraph. Our algorithm is able to handle many more identically shaped
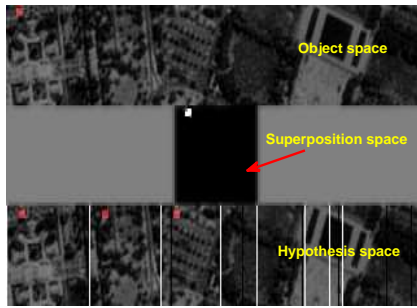
Fig. 11. (Media 2) showing the decoding of 4 targets that appear at random times and for random durations at random locations. The movie shows the successful decoding of all the 4 targets. Each frame shows the encoded object space ($\delta = \delta_{max}$), the corresponding measured superposition space (with static background subtracted out) and the resulting hypothesis space. (Movie size: 1.25 MB)

targets than the four chosen here, but using only a few targets allows the reader to easily follow the decoding logic shown in the movie.

The movie in Fig. 11 shows four identically shaped color-coded targets in the object space. Color coding allows easy target discrimination for the reader. The first target in the object space is the red target which is followed by a white target. Both these targets are decoded via missing ghosts logic as they travel through the object space. This can be seen in the hypothesis space where the ambiguity in the locations of these two targets is completely removed even before the targets reach a boundary. The green target appears next and is followed by the blue target. Since the blue target appears in a region of overlap of $fov_1$ and $fov_2$, it is almost immediately decoded. The green target is decoded when it reaches the region of overlap between $fov_1$ and $fov_2$. Thus all four targets are successfully decoded.

The movie shows us that the decoding time of targets differs depending on the target location and the target velocity. Also, because of possible errors in measurement it is possible that we incorrectly decode a target. This is especially true for targets with very low SNR. Therefore, decoding time and measurement errors can affect system performance. We next consider two metrics useful in quantifying this system performance. One metric considers the effect of shift resolution on average decoding time and the other considers the probability of incorrect decoding. We investigate these metrics in the following sub-sections.

## 5.2. *Average decoding time and area coverage efficiency ($\eta$)*

A small shift resolution (small $\alpha$) implies that the degree of overlap between adjacent sub-FOVs is small. One potential disadvantage of small shift resolution is that the average time it takes to decode a target (decoding time) increases.

The problem can arise as follows: when a new target ghost appears in superposition space, velocity estimates are not instantaneously available. Therefore, it takes some time to determine if the new ghost should be associated with an existing group, or if it is due to a new target altogether. To get velocity information we must wait a few time steps while the Kalman tracker updates the state vector and obtains a stable velocity estimate. This waiting period can present a problem, especially when the target is in a sub-FOV with a small overlap. If the target has a high $x$-velocity, it is possible that the target may not stay in the overlap region for enough time for the Kalman tracker to ascertain the target velocity. This results in increased decoding time.

Moreoever, for systems with small shift resolution, the distance between two different overlap regions is relatively large (equivalent to lower boundary density), which again tends to increase decoding time.
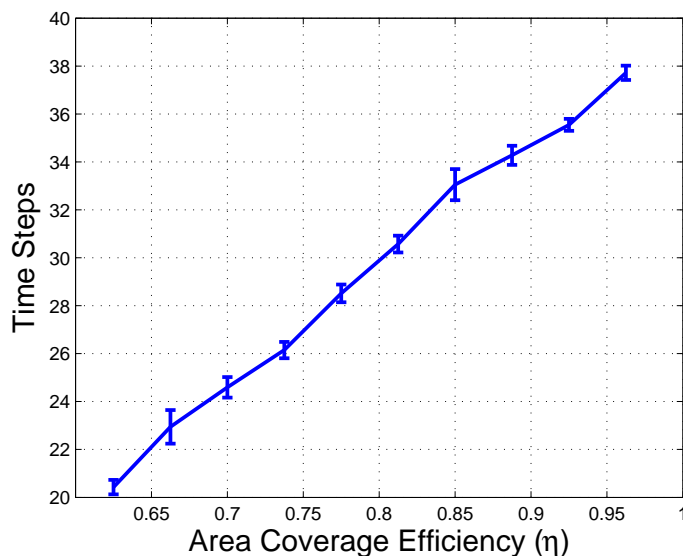


Fig. 12. Plot of decoding time as a function of the area coverage efficiency $\eta$. The plot shows the error bars representing the $\pm 1$ standard deviation of the decoding time from the mean.

A large shift resolution ($\alpha$ close to 1), on the other hand, suffers less from the above mentioned disadvantages, but has a smaller area coverage due to larger overlaps. Hence, shift resolution controls the trade-off between decoding time and area coverage efficiency. A small shift resolution provides larger area coverage, but larger decoding times. A large shift resolution provides smaller area coverage, but shorter decoding times. We quantify this result by plotting the decoding time as a function of the area coverage efficiency in Fig. 12. The plot also shows the error bar representing $\pm 1$ standard deviation of the decoding time from the mean. The plot was computed by averaging the decoding times of 300 targets which passed through the object space in batches of size uniformly distributed between 5 through 10. In each batch the targets appeared at random times with random velocities and at random locations, and for random durations in a manner identical to the targets in the movie in Fig. 11. The plot shows an approximately linear relationship between the two metrics. This is expected because displacement and time are linearly related through velocity. An increase in overlap due to a larger shift resolution for the same target with the same target velocity decreases the amount of time it takes for the target to reach the regions of overlap because these regions are now wider and cover more area. The reduced decoding time is directly related to this reduced travel time to reach the regions of overlap. The reasoning is the same when using "missing ghosts" logic to decode targets. The larger the overlaps, the faster we can reduce the ambiguity between hypothesized targets, and the smaller the decoding time. It is important to note here that complete decoding using "missing ghosts" plays a less prominent role in affecting the decoding time because it can completely decode only those targets which are in the first sub-FOV from the left. For all other cases the absence of shifts reduces, but does not completely eliminate, all the target ghosts and as a result does not affect decoding time.

### 5.3. Probability of decoding error

We next consider the probability of decoding error, which is defined as the probability that the target pattern decoded from superposition space measurements is an incorrect pattern. In the presence of noise and other distortion, the estimated position of a target ghost in superposition space will be subject to error. Therefore, the difference between two ghost positions, which is the criterion for target decoding in a shift-encoded system, will also be subject to error. These errors can lead to the wrong pattern being detected, which will cause the target to be decoded to the wrong location. Furthermore, as the shift resolution $\delta$ is decreased, more fidelity in estimating target shifts is required.

Figure 13 shows the results from a simplified calculation of decoding error for a single target present in a region of overlap between $M$ sub-FOVs. We first consider overlaps between two sub-FOVs and then extend the result to overlaps between three and four sub-FOVs. We assume, as we did for the simulation example in sub-section 5.1, that the imaging system superimposes 8 sub-FOVs onto a single FPA. The width of each sub-FOV is $500\Delta r$ distance units, where $\Delta r$ is the object space resolution. We assume that the error in estimating the position of a ghost in superposition space is Gaussian distributed with variance determined by the Cramer-Rao Bound (CRB) [24] applicable to this problem. The CRB is

$$var[\hat{x}] \geq \frac{1}{\text{SNR} \cdot B_{rms}}, \tag{6}$$

where SNR is proportional to the target intensity and $B_{rms}$ is the root-mean-square (rms) bandwidth of the target's intensity profile in the encoded dimension. For example, we have used a symmetric triangular intensity pattern, which has a closed-form rms bandwidth of $12/W_t^2$ [pixels$^{-2}$] where $W_t$ is the target width in pixels.

We first consider the case of a single target present in a region overlapped by two sub-FOVs ($M = 2$). The target pattern for this case consists of a single distance between two ghosts. If the position errors on the two ghosts are independent, then the variance of the distance estimate is twice the variance in (6). We assume that the target is decoded only if the measured overlap matches an allowed overlap from the overlap set $\mathbf{O}$ to within some prescribed tolerance $\varepsilon$. (Note that the distance between the two ghosts is related to the overlap through $S_i = W_{fov} - O_i, i = 0, 1, \ldots, N_s - 2$. See section 4 for more details.) If the measured overlap is not within $\varepsilon$ of a valid overlap then the target remains undecoded. For example, if the true shift is $m\delta$, a decoding error is made only if the measured overlap $\hat{d}$ satisfies $(m+k)\delta - \varepsilon \leq \hat{d} \leq (m+k)\delta + \varepsilon$ where $k$ is a non-zero integer and $(m+k)\delta$ is a valid overlap. The probability of decoding error can now be computed by integrating the Gaussian error probability distribution over the error region. This error probability is conditioned on $m\delta$ being the true overlap. Therefore, to calculate the total probability of decoding error we have to know the *a priori* probability of the true overlap being $m\delta$. We assume that this probability is uniformly distributed. The value of $\varepsilon$, in general, is dependent on the structure of the overlap set $\mathbf{O}$. We, however, have chosen the overlaps to be multiples of the shift resolution $\delta$. As a result all the overlap values are equally spaced from the adjacent ones. We can therefore let $\varepsilon$ be some fixed tolerance value less than or equal to $\delta/2$, where $\delta$, the shift resolution, is the separation between two successive overlaps. When $\varepsilon = \delta/2$ we always make a decoding decision. If $\varepsilon < \delta/2$, there are cases where the measured overlap does not lie within the tolerance limit of any element of $\mathbf{O}$ and we let the target remain undecoded. In contrast, if the overlap set contains random but unique overlaps, the tolerance $\varepsilon$ is a function of $\mathbf{O}$. For instance, the $\varepsilon$ tolerance value for overlaps with a large spacing between them will be different from the $\varepsilon$ tolerance value for the overlaps that are closely spaced, especially for the case where we always make a decoding decision. The tolerance value, therefore, will have to be adjusted according to the overlaps.
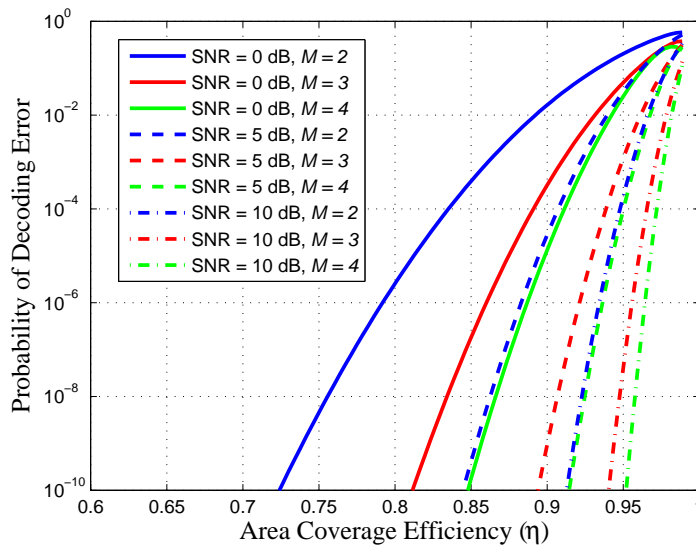
Fig. 13. Plot of decoding error versus area coverage efficiency for different SNRs and different values of $M$, the number of sub-FOVs that overlap.

We can extend the above result to the general case where $M$ sub-FOVs overlap. In our example we can have a maximum of $M = 4$. For the case of $M = 3$, instead of measuring a single overlap $\hat{d}$, we measure two overlaps $\hat{d}_1$ and $\hat{d}_2$ resulting from the pair-wise distances between three target ghosts in the superposition space. Therefore, we now have a 2-D Gaussian error probability distribution. The probability of decoding error is calculated by integrating this 2-D distribution over the region given by $(m_1 + k_1)\delta - \varepsilon \leq \hat{d}_1 \leq (m_1 + k_1)\delta + \varepsilon$ and $(m_2 + k_2)\delta - \varepsilon \leq \hat{d}_2 \leq (m_2 + k_2)\delta + \varepsilon$. Here $m_1\delta$ and $m_2\delta$ are the true overlaps, and $k_1$ and $k_2$ are non-zero shifts such that $(m_1 + k_1)\delta$ and $(m_2 + k_2)\delta$ are valid overlaps. We again assume that the probability of the true overlaps being $m_1\delta$ and $m_2\delta$ is uniformly distributed. Extention to $M = 4$, where we have a 3-D Gaussian error probability distribution is straightforward.

Figure 13 shows the probability of decoding error versus area coverage efficiency for $\varepsilon = \delta/2, W_t = 10\Delta r$ distance units, and different values of SNR and $M$. As the shift resolution decreases, area coverage efficiency increases, but so does probability of decoding error. Thus, the choice of shift resolution is a compromise between the area coverage and the probability of incorrectly decoding the target location. We also observe that for fixed SNR, as we increase $M$ the decoding error decreases. We therefore conclude that longer target patterns make the decoding process more robust and less prone to decoding errors.

### 5.4. Experimental results

To illustrate how 1-D spatial shift encoding of multiple sub-FOVs can be performed in real-world applications we conducted an experiment using the optical setup proposed in Fig. 3(b). The object space used for the experiment was an aerial map of the Duke University campus, and a laser pointer was moved across it during the video aquisition to simulate a single moving target. The object space was 24-mm high and 162-mm wide, and was imaged using the multiplexer in Fig. 3(b) onto a commercial video camera (SONY, DCR-SR42). By adjusting the tilts of the mirrors shown in Fig. 3(b), we obtained an overlap set $\mathbf{O} = \{3, 7, 11, 20, 14, 24, 28\} \times 1\text{mm}$ which deviated slightly from the ideal scenario of $\{0, 5, 10, 20, 15, 25, 30\} \times 1\text{mm}$. In building

the setup, care was taken to make the path lengths travelled by light from each sub-FOV close to equal. However, there was a slight difference in path lengths which resulted in varying magnification of some sub-FOVs. Therefore, the size of each sub-FOV was not uniform: $\mathbf{W}_{fov} = \{35, 34, 33, 35, 33, 32, 33, 34\} \times 1\text{mm}$ and $\mathbf{H}_{fov} = \{24, 23, 22, 23, 22, 22, 22, 23\} \times 1\text{mm}$, where the $i^{\text{th}}$ elements of $\mathbf{W}_{fov}$ and $\mathbf{H}_{fov}$ are the width and height of $fov_i$, respectively.
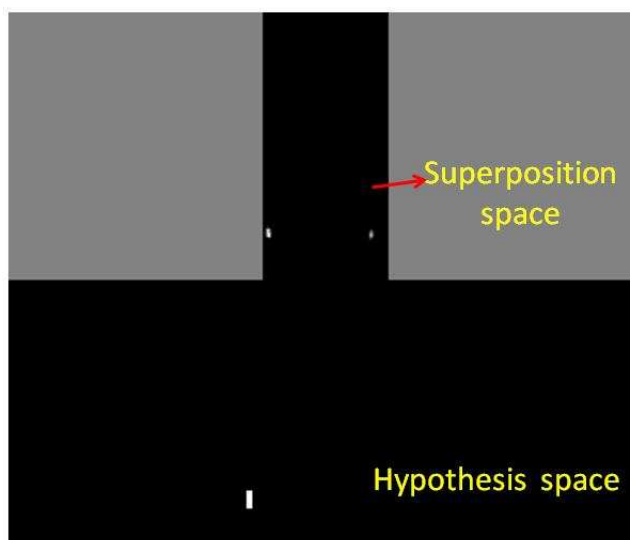


Fig. 14. Experimental data movie showing successful decoding a target moving through the object space. Intial target ambiguity is reduced using "missing ghosts" logic and is completely removed as the target enters the region of overlap. ([Media 3](#)) (Movie size: 3.9 MB)

Figure 14 is a movie we made using this experimental setup. The movie shows the measured superposition space along with the corresponding hypothesis space. Using the decoding logic discussed in Section 4 we are able to decode the moving target as it enters the region of overlap. The movie also shows how the "missing ghosts" logic reduces ambiguity about the target's true location in the hypothesis space. The small deviations of the overlaps from their true values do not affect the performance because all the overlaps are still unique. Uniqueness of overlaps is both the necessary and sufficient condition for the applicability of our decoding strategy. Also, the slight variations in magnification of the sub-FOVs do not affect the decoding performance.

## 6. Conclusion

In this paper we proposed a novel technique to track targets without conventional image reconstructions. The method is based on optical multiplexing of encoded sub-FOVs to create superposition space data that can be used to decode target positions in object space. We proposed a class of low complexity multiplexed imagers to perform optical encoding and showed that they are light, cheap and have a simple design in comparison to the conventional imagers. We discussed different encoding schemes based on spatial shifts, rotations, and magnification with special emphasis on 1-D spatial shift encoding. We showed, based on both simulation and experimental data, that the proposed method does indeed localize targets in object space and provides continuous target tracking capability. We have also studied the trade-offs between area coverage efficiency, compression ratio, and decoding time, and decoding error as a function of shift resolution and SNR.