

ORIGINAL PAPER

Multi-view in lensless compressive imaging

HONG JIANG, GANG HUANG AND PAUL WILFORD

Multi-view images are acquired by a lensless compressive imaging architecture, which consists of an aperture assembly and multiple sensors. The aperture assembly consists of a two-dimensional array of aperture elements whose transmittance can be individually controlled to implement a compressive sensing matrix. For each transmittance pattern of the aperture assembly, each of the sensors takes a measurement. The measurement vectors from the multiple sensors represent multi-view images of the same scene. We present theoretical framework for multi-view reconstruction and experimental results for enhancing quality of image using compressive measurements from multiple sensors.

Keywords: Lensless compressive imaging, Multi-view imaging, Image reconstruction

Received 30 May 2014; Revised 10 November 2014; Accepted 11 November 2014

1. INTRODUCTION

Lensless compressive imaging [1] is an effective architecture to acquire images using compressive sensing [2]. The architecture, which is illustrated in Fig. 1, consists of an aperture assembly and a sensor of a single detecting element; no lens is used. The transmittance of each aperture element in the aperture assembly is individually controllable. The sensor is used to take compressive measurements. A sensing matrix is implemented by adjusting the transmittance of the individual aperture elements according to the values of the sensing matrix. This architecture is distinctive in that the images acquired are not formed by any physical mechanism, such as a lens [3, 4] or a pinhole [5]. There are no aberrations introduced by a lens such as blurring due to scenes out of focus. Furthermore, the same architecture can be used for acquiring multimodal signals such as infrared, Terahertz [6, 7], and millimeter wave images [8]. This architecture has application in surveillance [9].

The lensless compressive imaging architecture is well-suited for multi-view imaging because multiple sensors may be used in conjunction with one aperture assembly as shown in Fig. 2. The cost of obtaining an additional viewpoint is simply that of adding a sensor to the device. For a given setting of transmittance, each sensor takes a measurement, and therefore, for a given sensing matrix, the sensors produce a set of measurement vectors simultaneously. Each measurement vector can be used to reconstruct an image independently without taking into consideration of other measurement vectors. However, although the images

from different sensors are different, there is a high correlation between them, especially when the sensors are located closely to one another and when the scene is far away. The correlation between the images can be exploited to enhance the quality of the reconstructed images.

Multiple sensors with one aperture assembly may be used in the following three ways:

Multi-view. The measurement vectors from multiple sensors represent images of different views of a scene, creating multi-view images, especially when the scene is near. This architecture allows a simple device to capture multi-view images simultaneously.

Measurement increase. When the scene is sufficiently far away, the measurement vectors from the sensors may be considered to be independent measurements of a common image and they may be concatenated into a larger set of measurements to reconstruct the common image. This effectively increases number of measurements that are taken for the image in a given duration of time.

Higher resolution. When the scene is sufficiently far away, and when the sensors are properly positioned, the measurement vectors from the sensors may be considered to be the measurements made from a higher resolution image, and they may be used to reconstruct an image of the higher resolution than the number of aperture elements.

A) Main contributions of this paper

The purpose of this paper is twofold. First, we present a theoretical framework for reconstructing multi-view images using joint reconstruction, which exploits the correlation between the multiple viewpoints. Secondly, we present experimental results to demonstrate how the multiple sensors can be used in each of the above three ways.

Bell Labs, Alcatel-Lucent, Murray Hill, NJ 07974, USA

Corresponding author:

H. Jiang

Email: hong.jiang@alcatel-lucent.com

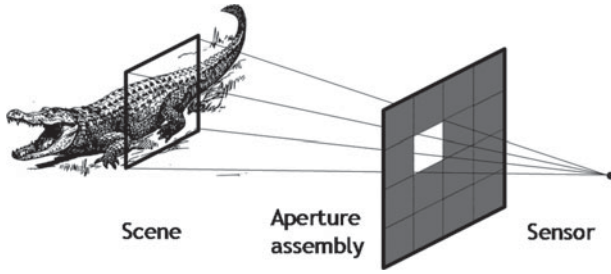


Fig. 1. Lensless compressive imaging architecture.

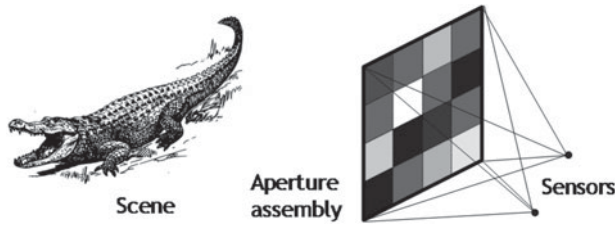


Fig. 2. Lensless compressive imaging with two sensors.

B) Related work

The lensless compressive imaging was first proposed in [1], which is related, but quite different from the work in [3–5]. Two lenses are used to accomplish compressive imaging in [3, 4]. While the architecture of [5] does not use a lens, it does not compress image during acquisition, as [1] and [3, 4] do. Furthermore, in the architecture of [3–5], images are physically formed before they are digitized and acquired.

The discussion in this paper is mainly focused on creating one image using measurements from multiple sensors, although the measurements can also be used for 3D imaging and depth maps, in a way similar to those of [10–12].

The coded aperture [13, 14] is a special case of the multi-view of lensless compressive imaging in which a very large number of sensors are used and only one measurement is taken from each sensor.

C) Organization of the paper

In the next section, the lensless compressive imaging is reviewed, which provides necessary background for multi-view. In Section III, multi-view imaging is discussed. In particular, how the measurements from multiple sensors can be used in a joint reconstruction is described. Experiments and simulations are given in Section IV, followed by Conclusion in Section V.

II. LENSLESS COMPRESSIVE IMAGING

In this section, we formally define what an image is in the lensless compressive imaging and how it is related to the measurements from the sensor. In particular, we will describe how a digitized image can be reconstructed from the measurements taken from the sensor.

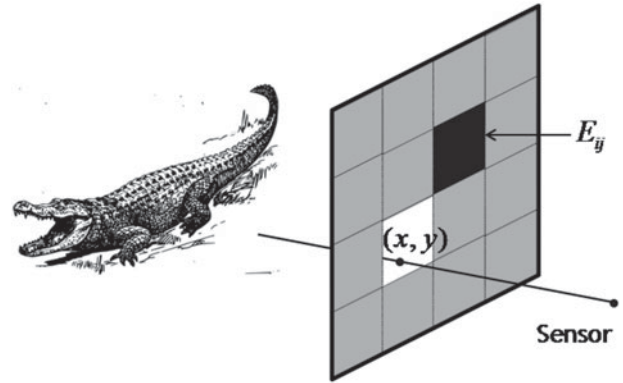


Fig. 3. A ray is defined for each point on the region of aperture assembly.

A) Virtual image

Let the aperture assembly be a rectangular region on a plane with $(x, y) \in \mathbb{R}^2$ coordinate system. For each point (x, y) on the aperture assembly, there is a ray starting from a point on the scene, passing through the point (x, y) , and ending at the sensor, as shown in Fig. 3. Therefore, there is a unique ray associated with each point (x, y) on the aperture assembly, and its intensity arriving at the aperture assembly at time t is denoted by $r(x, y; t)$. Then an image $I(x, y)$ of the scene is defined as the integration of the ray in a time interval Δt :

$$I(x, y) = \int_0^{\Delta t} r(x, y; t) dt. \tag{1}$$

Note that the image in (1) is defined mathematically, and although it is defined in the region of the aperture assembly, there is not an actual image physically formed in the lensless compressive imaging architecture. For this reason, the image of (1) is said to be a virtual image. A virtual image $I(x, y)$ can be considered as an analog image because it is continuously defined in the region of the aperture assembly.

Let the transmittance of the aperture assembly be defined as $T(x, y)$. A measurement made by the sensor is the integration of the rays through the aperture assembly modulated by the transmittance, and it is given by

$$\bar{z}_T = \iint T(x, y) I(x, y) dx dy. \tag{2}$$

Although the virtual image discussed above is defined on the plane of the aperture assembly, it is not necessary to do so. The virtual image may be defined on any plane that is placed in between the sensor and the aperture assembly and parallel to the aperture assembly.

B) Digitized image

The virtual image defined by (1) can be digitized by the aperture assembly. For any given region R on the plane of the aperture assembly, we define its characteristic function as

$$\mathbf{1}_R(x, y) = \begin{cases} 1, & (x, y) \in R \\ 0 & (x, y) \notin R \end{cases}. \tag{3}$$

Let the aperture elements of the aperture assembly be indexed by $(i, j) \in \mathbb{Z}^2$. We denote by E_{ij} the region on the plane of the aperture assembly defined by the (i, j) th aperture element as shown in Fig. 3. Each element E_{ij} can be used to define a pixel in the digitized image and the pixel value of the image at pixel (i, j) is the integration of the rays passing through the region E_{ij} and it is given by

$$I(i, j) = \iint_{E_{ij}} I(x, y) dx dy, \\ = \iint \mathbf{1}_{E_{ij}}(x, y) I(x, y) dx dy. \quad (4)$$

Note that we use $I(i, j)$ to denote a digitized image of a virtual image $I(x, y)$ which is considered to be analog.

Equation (4) defines the digitized image $I(i, j)$. In compressive sensing, it is often mathematically convenient to reorder a pixelized image which is a two-dimensional (2D) array into a one dimensional (1D) vector. Let q be a mapping from a 2D array to a 1D vector defined by, for example,

$$q : \mathbb{Z}^2 \rightarrow \mathbb{Z}, \quad \text{so that} \\ q(i, j) = i + (j - 1)N_r, \quad i, j = 1, 2, \dots, \quad (5)$$

where N_r is the number of rows in the aperture assembly. Then the pixelized image $I(i, j)$ can be represented as a vector whose components are I_n where $n = q(i, j)$. When there is no risk of confusion, we will simply use I to denote the pixelized image, either as a 2D, or a 1D vector, interchangeably. The total number of pixels in image I is denoted by N .

C) Compressive measurements and reconstruction

When the aperture assembly is programmed to implement a compressive sensing matrix, the transmittance of each aperture element is controlled according to the value of the corresponding entry in the sensing matrix. For the m th measurement, the entries in row m of the sensing matrix are used to program the transmittance of the aperture elements. Specifically, let sensing matrix $A \in \mathfrak{R}^{M \times N}$ be a random matrix whose entries, a_{mn} , are random numbers between 0 and 1. Then, for the m th measurement, the transmittance of the aperture assembly is given by

$$T_m(x, y) = \sum_{i,j} a_{m,q(i,j)} \mathbf{1}_{E_{ij}}(x, y). \quad (6)$$

Therefore, according to (2), the measurements are given by

$$z_m \triangleq \bar{z}_{T_m} = \iint T_m(x, y) I(x, y) dx dy, \\ = \sum_{i,j} a_{m,q(i,j)} \iint \mathbf{1}_{E_{ij}}(x, y) I(x, y) dx dy, \quad (7) \\ = \sum_{i,j} a_{m,q(i,j)} I(i, j), \quad m = 1, \dots, M.$$

Equation (7) is the familiar form of compressive measurements if the pixelized image $I(i, j)$ is reordered into a

vector by the mapping q . Indeed, in the vector form, (7) is tantamount to

$$z_m = \sum_{i,j} a_{m,q(i,j)} I(i, j) = \sum_n a_{mn} I_n, \quad \text{or} \\ z = A \cdot I. \quad (8)$$

In above, $z \in \mathfrak{R}^M$ is the measurement vector of length M , $A \in \mathfrak{R}^{M \times N}$ is the sensing matrix and $I \in \mathfrak{R}^N$ is the vector representation of the pixelized image $I(i, j)$.

It is well known [2, 9] that the pixelized image I can be reconstructed from the measurement vector z by, for example, solving the following minimization problem:

$$\min_{I \in \mathfrak{R}^N} \|W \cdot I\|_1, \quad \text{subject to} \quad A \cdot I = z, \quad (9)$$

where W is some sparsifying operator such as total variation [15] or framelets [9].

D) Angular dependence of incident rays

The definition of the virtual image $I(x, y)$ and its digitization $I(i, j)$ given above is theoretical, defined in terms of an ideal sensor which is, in particular, independent of the angle of incident rays from the scene. When a practical sensor is considered, the definition of the image I (meaning $I(x, y)$ or $I(i, j)$) does not change, but the measurement vector z of (7)–(9) would now depend on the property of the practical sensor. Our goal is still to compute, or approximate, the previously defined I , but now from the practically obtained measurement vector z .

Equations (7)–(9) hold only in the ideal case in which the response of the sensor is independent of the angle of incident rays from the scene. When the response of the sensor is a function of the angle of the incident rays, the measurement vector z is given by an equation different from those in (7)–(9) to be discussed below.

For a fixed geometry and relative locations of the aperture assembly and the sensor, the angle of incidence of a ray from the scene to the sensor is a function of (x, y) which is the position where the ray intersects the aperture assembly, see Fig. 3. Therefore, the response of the sensor is also a function of (x, y) . Define the response of the sensor as $\bar{G}(x, y)$. Then the measurement vector z of (7), when the effect of angular dependence is considered, becomes

$$z_m = \iint T_m(x, y) \bar{G}(x, y) I(x, y) dx dy, \\ = \sum_{i,j} a_{m,q(i,j)} \iint \mathbf{1}_{E_{ij}}(x, y) \bar{G}(x, y) I(x, y) dx dy, \\ \approx \sum_n a_{mn} g_n I_n, \quad \text{where} \\ g_n = \iint \mathbf{1}_{E_{q(i,j)}}(x, y) \bar{G}(x, y) dx dy. \quad (10)$$

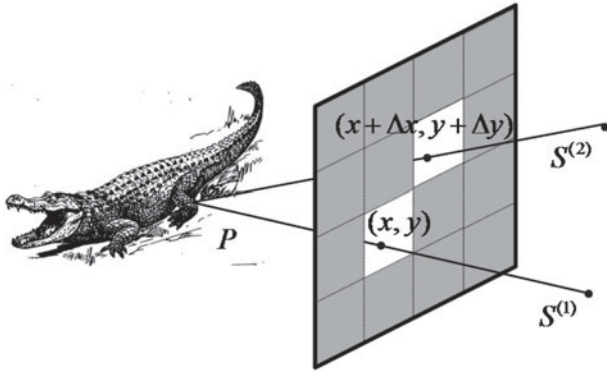


Fig. 4. Multiple sensors are used with one aperture assembly to make multi-view images.

Consequently, when the effect of angular dependence is considered, (9) becomes

$$\min_{I \in \mathbb{R}^N} \|W \cdot I\|_1, \quad \text{subject to } A \cdot G \cdot I = z,$$

$$\text{where } G = \begin{bmatrix} g_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & g_N \end{bmatrix}. \quad (11)$$

Equation (11) differs from (9) by a diagonal matrix G , which reflects the angular dependence of incident rays of the sensor. The diagonal matrix G only depends on the property of the sensor and the location of the sensor relative to the aperture assembly. In particular, different location of the sensor results in a different G . In practice, the entries of G for each sensor can be approximated in a calibration process. For example, in order to find $g_n, n = 1, \dots, N$, for a sensor, a uniform white background is used as the scene, and N measurements are taken by the sensor when the aperture elements are opened and closed one by one. Each measurement is then an approximation of the corresponding g_n .

III. MULTI-VIEW IMAGING

Multiple sensors may be used in conjunction with one aperture assembly as shown in Fig. 4, in which two sensors are drawn. A virtual image can be defined for each of the sensors; for example, $I^{(k)}(x, y)$ is the virtual image associated with sensor $S^{(k)}$, where the superscript k is used for indexing the multiple sensors. These images are multi-view images of a same scene.

For a given pattern of transmittance $T(x, y)$, each sensor takes a measurement, and therefore, for a given sensing matrix, the sensors produce a set of measurement vectors, $z^{(k)}$, simultaneously. Each measurement vector $z^{(k)}$ can be used to reconstruct a pixelized image $I^{(k)}$ by solving problem (9) independently without taking into consideration of other measurement vectors. However, although the images $I^{(k)}$ are different, there is a high correlation between them, especially when the sensors are located near one another and when the scene is far away. The correlation between

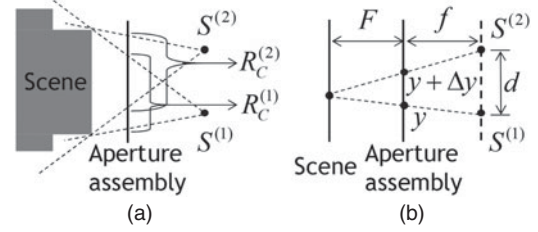


Fig. 5. Various definitions for two sensors on a plane parallel to the plane of aperture assembly. The illustration is made on a plane perpendicular to the plane of aperture assembly so that the aperture assembly is illustrated as a vertical line. The vertical line marked by “aperture assembly” illustrates the plane on which the virtual image is defined. In (A), the regions marked by $R_C^{(1)}$ and $R_C^{(2)}$ illustrates the common regions of $I^{(1)}$ and $I^{(2)}$, respectively. The regions $R_D^{(1)}$ and $R_D^{(2)}$ are not marked, but they are the regions on the vertical line that are complements of $R_C^{(1)}$ and $R_C^{(2)}$, respectively.

the images can be exploited to enhance the quality of the reconstructed images.

In the rest of this section, we analyze the relationship between the images taken from two sensors. We will first establish a mapping between the common scenes in the two images for any scene geometry. This mapping makes it theoretically possible for the common scenes to be reconstructed jointly with measurements from both sensors. The mapping is, in general, not known explicitly if the scene geometry is unknown. However, when the scene is planar and parallel to the plane of aperture assembly, or when the scene is far away, the mapping can be simplified and explicitly formulated, and the joint reconstruction of the common scenes is made practical.

A) Image decomposition

We consider two sensors, $S^{(1)}$ and $S^{(2)}$, that are placed on a same plane parallel to the plane of aperture assembly, as shown in Fig. 5. The sensors define two virtual images $I^{(1)}(x, y)$ and $I^{(2)}(x, y)$. We want to explore common component between them.

The area of the aperture assembly can be divided into two disjoint regions, $R_C^{(1)}$ and $R_D^{(1)}$, according to $S^{(1)}$. In the simplest term, $R_C^{(1)}$ consists of the scene that can be also seen by $S^{(2)}$; that is, the objects appearing in $R_C^{(1)}$ are common in both images, $I^{(1)}(x, y)$ and $I^{(2)}(x, y)$. $R_D^{(1)}$ consists of the scene that can be only seen by $S^{(1)}$; that is, the objects appearing in $R_D^{(1)}$ can only be found in $I^{(1)}(x, y)$. The definition of the two regions can be made more precise using the rays from the two sensors.

As shown in Fig. 4, any point (x, y) on the aperture assembly defines a ray that starts from the sensor $S^{(1)}$ and passes through (x, y) . The ray must ends at a point P in the scene. Now if a ray emitted from point P can reach the sensor $S^{(2)}$ through the aperture assembly without obstruction by other objects of the scene (with all aperture elements open), then $(x, y) \in R_C^{(1)}$. Otherwise, if no rays from P can reach the sensor $S^{(2)}$ (with all aperture elements open), then $(x, y) \in R_D^{(1)}$. $R_C^{(2)}$ and $R_D^{(2)}$ can be similarly defined as above

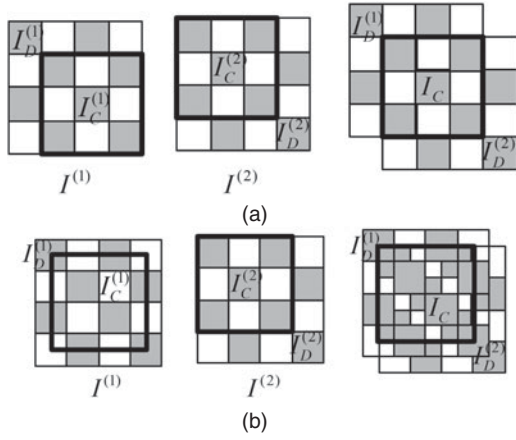


Fig. 6. Decomposition of images of a planar scene from two sensors. $I^{(k)} = I_C^{(k)} + I_D^{(k)}, k = 1, 2$. $I_C^{(1)}$ and $I_C^{(2)}$ are shifts of the common image, I_C . (a) The sensor distance is an integer multiple of the size of the aperture elements. (b) The sensor distance is a non-integer multiple of the size of the aperture elements.

by reversing the role of $S^{(1)}$ and $S^{(2)}$. $R_C^{(1)}$ and $R_C^{(2)}$ are illustrated in Fig. 5(A) in 1D view.

Incidentally, the definition of $R_C^{(1)}$ and $R_C^{(2)}$ also defines a one-to-one mapping between them. The points where the rays $\overline{PS^{(1)}}$ and $\overline{PS^{(2)}}$ intersect the aperture assembly are mapped into each other. The mapping is defined as

$$\begin{aligned}
 U^{12} : R_C^{(1)} &\rightarrow R_C^{(2)} \text{ so that } U^{12}(x, y) = (x + \Delta x, y + \Delta y), \\
 U^{21} : R_C^{(2)} &\rightarrow R_C^{(1)} \text{ so that } U^{21}(x + \Delta x, y + \Delta y) = (x, y),
 \end{aligned}
 \tag{12}$$

where the relationship between points (x, y) and $(x + \Delta x, y + \Delta y)$ is shown in Fig. 4.

Now the virtual images $I^{(k)}(x, y)$ can be decomposed using the characteristic functions of $R_C^{(k)}$ and $R_D^{(k)}$ as follows

$$\begin{aligned}
 I^{(k)}(x, y) &= I_C^{(k)}(x, y) + I_D^{(k)}(x, y) \\
 I_C^{(k)}(x, y) &= I^{(k)}(x, y) \mathbf{1}_{R_C^{(k)}}(x, y) \quad , \quad k = 1, 2. \\
 I_D^{(k)}(x, y) &= I^{(k)}(x, y) \mathbf{1}_{R_D^{(k)}}(x, y)
 \end{aligned}
 \tag{13}$$

Furthermore, $I_C^{(1)}(x, y)$ and $I_C^{(2)}(x, y)$ are related through the following equations:

$$\begin{aligned}
 I_C^{(2)}(x, y) &= I_C^{(1)}(U^{21}(x, y)), \\
 I_C^{(1)}(x, y) &= I_C^{(2)}(U^{12}(x, y)).
 \end{aligned}
 \tag{14}$$

The decomposition, $I_C^{(k)}(x, y)$ and $I_D^{(k)}(x, y), k = 1, 2$, is illustrated in Fig. 6 for planar scenes. The relation of Fig. 6 holds because in lensless compressive imaging, the three planes, the scene plane, the plane of aperture assembly and the plane on which the sensors are located, are parallel to one another, as illustrated in Fig. 5(b).

The significance of the decomposition (13) is that the two virtual images are decomposed into three components: one component is common to both images, and the other

two components are unique to each individual image. More specifically, if we define the common component as

$$I_C(x, y) = I_C^{(1)}(x, y), \tag{15}$$

then we have

$$\begin{aligned}
 I^{(1)}(x, y) &= I_C(x, y) + I_D^{(1)}(x, y), \\
 I^{(2)}(x, y) &= I_C(U^{21}(x, y)) + I_D^{(2)}(x, y).
 \end{aligned}
 \tag{16}$$

Since $I_C(x, y)$ is common in both images, its reconstruction may make use of the measurements from both sensors, and therefore, its quality may be enhanced as compared to only one sensor is used.

B) Joint reconstruction

The components of the virtual images, $I_C(x, y), I_D^{(1)}(x, y)$ and $I_D^{(2)}(x, y)$, can be pixelized to get three vector components $I_C, I_D^{(1)}$ and $I_D^{(2)}$. Referring to Fig. 6, the decomposition is similar to (16) and given by

$$\begin{aligned}
 I^{(1)} &= I_C + I_D^{(1)}, \\
 I^{(2)} &= V \cdot I_C + I_D^{(2)}.
 \end{aligned}
 \tag{17}$$

In above, V is a matrix that performs shift and interpolating functions to approximate the operation of mapping U^{21} defined in (12), as given by

$$\begin{aligned}
 V : \mathfrak{R}^N &\rightarrow \mathfrak{R}^N, \text{ such that for all } I_C \in \mathfrak{R}^N, \\
 (V \cdot I_C)_{q(i,j)} &\approx \iint \mathbf{1}_{E_{ij}}(x, y) I_C(U^{21}(x, y)) dx dy.
 \end{aligned}
 \tag{18}$$

In other words, $V \cdot I_C$ is a vector that approximates the pixelized $I_C(U^{21}(x, y))$.

The vector components $I_C, I_D^{(1)}$ and $I_D^{(2)}$ may be jointly reconstructed from the two measurement vectors, $z^{(1)}$ and $z^{(2)}$, made from the two sensors,

$S^{(1)}$ and $S^{(2)}$. Let A be the sensing matrix with which the measurements $z_m^{(1)}$ and $z_m^{(2)}$ are made. Then the optimization problem to solve is

$$\begin{aligned}
 \min_{I_C \in \mathfrak{R}^N} & \|W \cdot I_C\|_1 + \frac{\sigma}{2} \sum_{k=1}^2 \left\| W \cdot I_D^{(k)} \right\|_1, \text{ subject to} \\
 A \cdot I_C + A \cdot I_D^{(1)} &= z^{(1)}, \\
 A \cdot V \cdot I_C + A \cdot I_D^{(2)} &= z^{(2)}.
 \end{aligned}
 \tag{19}$$

In (19), $\sigma > 0$ is a normalization constant to account for the areas of the four regions $R_C^{(k)}$ and $R_D^{(k)}, k = 1, 2$. The value of the joint reconstruction (19) lies in the fact that there are only three unknown components in (19) with two constraints (given by $z^{(1)}$ and $z^{(2)}$), as compared to four unknown components with two constraints if the images are reconstructed independently from (9). Typically, I_C has much more nonzero entries than that of $I_D^{(1)}$ and $I_D^{(2)}$, hence the number of unknowns is reduced by almost a half.

Equation (19) is valid for sensors that are independent of incident angle. In practice, when the sensors have angular dependence of incident rays, the constraints in (19) need to be modified according to (11). Let $G^{(1)}$ and $G^{(2)}$ be the diagonal matrices obtained from the calibration of the effect of angular dependence of sensors $S^{(1)}$ and $S^{(2)}$, respectively. Then, the constraints in (19) should be replaced by

$$\begin{aligned} A \cdot G^{(1)} \cdot I_C + A \cdot G^{(1)} \cdot I_D^{(1)} &= z^{(1)}, \\ A \cdot G^{(2)} \cdot V \cdot I_C + A \cdot G^{(2)} \cdot I_D^{(2)} &= z^{(2)}. \end{aligned} \quad (20)$$

In the rest of the paper, for simplicity, we omit the effect of angular dependence, with the understanding that the effect can be taken care of by (20).

In general, problem (19) is quite difficult to solve because the regions $R_C^{(k)}$ and $R_D^{(k)}$, $k = 1, 2$ are not known *a priori*, and they should be part of the solution. This general problem requires further study and is not addressed in this paper; instead, we resort to solving a simplified problem where the regions $R_C^{(k)}$ and $R_D^{(k)}$, $k = 1, 2$, can be computed prior to the joint reconstruction.

C) Planar scene

When the scene is on a plane parallel to, and with a known distance from, the plane of aperture assembly, it is possible to work out explicit formulas for the mappings U^{12} and U^{21} of (12). As shown in Fig. 5(b), let us define the distance between two sensors to be d , the distance between the plane of the sensors and the plane of aperture assembly to be f and the distance between the scene plane and the aperture assembly to be F . Then the mapping U^{12} is given by

$$\begin{aligned} U^{12}(x, y) &= (x + \Delta x, y + \Delta y), \\ \sqrt{\Delta x^2 + \Delta y^2} &= \frac{F}{f + F} d, \\ (\Delta x, \Delta y) &\propto \overrightarrow{S^{(1)} S^{(2)}}. \end{aligned} \quad (21)$$

The last line in (21) means that the two vectors have the same angle, or orientation, in their respective planes.

In general, when the scene is non-planar, equation (21) still holds, but F is no long a constant. It is rather a function of position, i.e., $F = F(x, y)$, and it is also scene dependent. However, for the scene that is sufficiently far away, F is large compared to f so that $(F/(f + F)) \approx 1$, and therefore, equation (21) becomes

$$\begin{aligned} U^{12}(x, y) &= (x + \Delta x, y + \Delta y), \\ \sqrt{\Delta x^2 + \Delta y^2} &\approx d, \\ (\Delta x, \Delta y) &\propto \overrightarrow{S^{(1)} S^{(2)}}. \end{aligned} \quad (22)$$

According to (22), when the scene is sufficiently far away, the virtual images from the two sensors are approximately the same, except for a shift of distance d . Therefore, the common region $R_C^{(k)}$ covers the entire aperture assembly except for a border of width d . Consequently, compared to

the common image I_C , the images $I_D^{(1)}$ and $I_D^{(2)}$ have small energy. This implies that problem (19) is mainly a problem for the single image I_C , while using two measurement vectors $z^{(1)}$ and $z^{(2)}$, twice as many measurements as when each of the images, $I^{(1)}$ and $I^{(2)}$, is reconstructed independently as in (9). For this reason, multiple sensors may be considered as taking independent measurements for a same image if the scene is sufficiently far away. This can be used as a mechanism to increase the number of measurements taken during a given time duration.

If the distance between two sensors, d , is equal to an integer multiple of the size of the aperture elements, as illustrated in Fig. 6(a), then matrix V in (19) is simply a shift matrix. In other words, the entries of V are zero except for the entries on an off-diagonal, which are equal to 1, as in

$$V = \begin{bmatrix} \ddots & & 1 & & \ddots \\ & \ddots & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \ddots \\ \ddots & & & & & \ddots \end{bmatrix}_{N \times N}. \quad (23)$$

D) High resolution

For sufficiently far away scenes, multiple sensors may also be used as a mechanism to improve the resolution of the common image I_C . If the distance d between two sensors is a non-integer multiple of the size of the aperture elements, then $I^{(1)}$ and $I^{(2)}$ can be considered as two down-sampled images of a higher resolution image, see Fig. 6(b). The joint reconstruction can therefore be used to create a higher resolution image.

Specifically, equation (16) can be rewritten as

$$\begin{aligned} I^{(1)}(x, y) &= I_C(x, y) + I_D^{(1)}(x, y), \\ I^{(2)}(x, y) &= I_C(x - \Delta x, y - \Delta y) + I_D^{(2)}(x, y). \end{aligned} \quad (24)$$

If the distance d between two sensors is a non-integer multiple of the size of the aperture elements, then there is no overlapping of grid points $(x - \Delta x, y - \Delta y)$ with the grid points (x, y) . Therefore, equation (24) shows that images $I^{(1)}$ and $I^{(2)}$ comprise different sampling of the same image I_C , i.e. $I^{(1)}$ samples I_C at points (x, y) , while $I^{(2)}$ samples I_C at points $(x - \Delta x, y - \Delta y)$. Consequently, the measurement vectors $z^{(1)}$ and $z^{(2)}$ can be used to reconstruct the image I_C at both grid points (x, y) and $(x - \Delta x, y - \Delta y)$. This results in an image I_C that has a higher resolution than given by the aperture elements. This is illustrated in Fig. 6(b).

E) Coded aperture imaging

The lensless compressive imaging with multiple sensors can also be used to implement the coded aperture imaging [13, 14], and therefore, the coded aperture imaging is merely a special case of the multi-view lensless compressive imaging of this paper, in which an array of sensors are used and each sensor makes only one measurement.

Indeed, let L be the total number of sensors in the multi-sensor architecture of Fig. 4, then the constraints in (19) can be written as

$$\begin{aligned} A \cdot I_C + A \cdot I_D^{(1)} &= z^{(1)}, \\ A \cdot V^{(k)} \cdot I_C + A \cdot I_D^{(k)} &= z^{(k)}, \quad k = 2, \dots, L. \end{aligned} \tag{25}$$

When the number of sensors L is large enough, e.g., when L is in the same order of magnitude as N , the total number of aperture elements in the aperture assembly, it is possible to reconstruct an image by using only one aperture pattern so that only one measurement is taken by each sensor. In this case, the sense matrix A in (25) has only one row, i.e.

$$A = [a_1, a_2, \dots, a_N]. \tag{26}$$

Thus, the measurements, with one measurement from each sensor, can be rewritten as

$$\begin{aligned} z^{(1)} &= [a_1, \dots, a_N] \cdot I_C + b^{(1)}, \\ z^{(k)} &= [a_1, \dots, a_N] \cdot V^{(k)} \cdot I_C + b^{(k)}, \quad k = 2, \dots, L, \end{aligned} \tag{27}$$

where

$$b^{(k)} = [a_1, \dots, a_N] \cdot I_D^{(k)}, \quad k = 1, \dots, L. \tag{28}$$

In the matrix form, (27) becomes

$$z \triangleq \begin{bmatrix} z^{(1)} \\ \vdots \\ z^{(L)} \end{bmatrix} = C \cdot I_C + b, \tag{29}$$

where z is a column vector, each component of which is the measurement from one of the L sensors, the matrix C is a cyclic matrix whose first row is $[a_1, a_2, \dots, a_N]$ (with zeros padded) and the other rows are cyclic shifts of the first row. Vector b includes measurement noise as well as portion of the image that cannot be seen by all sensors. Equation (29) is the recognizable equation for the coded aperture imaging [14].

The analysis above not only demonstrates that the coded aperture imaging is a special case of the lensless compressive sensing, as shown in (29), but also it suggests an algorithm of reconstruction for the coded aperture imaging, which is to minimize the cost function in (19) subject to the constraints (29).

IV. EXPERIMENT

A) Prototype

A lensless compressive imaging prototype with two sensors is shown in Fig. 7. It consists of a transparent monochrome liquid crystal display (LCD) screen and two photovoltaic sensors enclosed in a light tight box. The LCD screen functions as the aperture assembly while the photovoltaic sensors measure the light intensity. The photovoltaic sensors



Fig. 7. Prototype device. Top: lab setup. Bottom left: the LCD screen as the aperture assembly. Bottom right: the sensor board with two sensors, indicated by the red circle.

are tricolor sensors, which output the intensity of red, green and blue lights.

The LCD panel is configured to display $302 \times 217 = 65\,534$ black or white squares. Each square represents an aperture element with transmittance of a 0 (black) or 1 (white). A Hadamard matrix of order $N = 65\,536$ is used as sensing matrix, which allows a total number of 65 534, corresponding to the total number of pixels in the image, independent measurements to be made by each sensor. In our experiments, we only make a fractional of the total number of measurements. We express the number of measurements taken and used in reconstruction as a percentage of the total number of pixels. For example, 25% of measurements means 16 384 measurements are taken and used in reconstruction, which is a quarter of the total number of pixels, 65 534. In each experiment, a set of measurements is obtained by each sensor simultaneously. The two sensors are placed such that there is almost no vertical offset, and there is a horizontal offset of approximately 3.5 pixels.

B) Multi-view

Measurements taken from each of the two sensors are used to reconstruct an image using (9), independent of the measurements from the other sensor. The total variation [15] is used for the sparsifying operator W . The results are two images that can be used for 3D display of a scene, as shown in Fig. 8. The left and right images in Fig. 8 are placed so that the cross-eyed view results in the correct stereogram.

Each of the images in Fig. 8 was reconstructed using 50% of measurements.



Fig. 8. Multi-view images. Left: image reconstructed from measurements of sensor 1. Right: image reconstructed from measurements of sensor 2.



Fig. 9. Reconstruction using measurements from two sensors.

C) Measurement increase

We compare the quality of images by individual and joint reconstructions in Fig. 9, which is composed of six images, arranged in two columns and three rows. On the top row, the two images are reconstructed by (9) using 12.5% (left) and 25% (right) of measurements taken from sensor 1 only. In the middle row, two images are the same; it is reconstructed by (19) using 12.5% of measurements from each of the two sensors (for a combined 25%). On the bottom row, the two images are reconstructed by (9) using 12.5% (left) and 25% (right) of measurements taken from sensor 2 only. We can make a couple of observations from Fig. 9. First, as expected, the images using 25% measurements from one sensor only are clearly better than the images using 12.5% measurements from one sensor only. That is, an image on the right column, top or bottom row, is better than an image on the left column, top or bottom row. Second, the image from joint reconstruction using measurements from

both sensors is better than images using 12.5% measurements from one sensor only, and as good as the images using 25% measurements from one sensor only, i.e., in the left column, the middle image is better than top and bottom; in the right column, all three images are similar. In reconstructing the image in the middle row, although a total of 25% of measurements are used, these measurements are taken in a time interval during which each sensor only takes 12.5% of measurements.

D) Higher resolution

In Fig. 10, the top and bottom images are reconstructed individually by (9) using 25% of measurements taken from each of sensor 1 and sensor 2, respectively. The middle image is reconstructed using joint reconstruction to a higher resolution, 604×217 , using 25% measurements from each of two sensors, taking the advantage that there is a 3.5 pixels horizontal offset between the two sensors. It is evident that the



Fig. 10. Reconstruction to higher resolution using measurements from two sensors.

image in the middle is sharper due to twice the horizontal resolution.

V. CONCLUSION

Lensless compressive imaging is an effective architecture to acquire multi-view images. The cost of obtaining an additional viewpoint is simply the cost of a single photodetector. The compressive measurements from multiple sensors in lensless compressive imaging may be used for multi-viewing, enhancing quality of an image by increasing number of measurements, or by increasing the resolution of the image.

ACKNOWLEDGEMENTS

The authors wish to thank Kim Matthews for continuing interest in, and insightful discussion and support on, the lensless compressive imaging architecture.

REFERENCES

- [1] Huang, G.; Jiang, H.; Matthews, K.; Wilford, P.: Lensless imaging by compressive sensing, in *IEEE Int. Conf. Image Processing, ICIP 2013*, August 2013, 1–5.
- [2] Candès, E.; Romberg, J.; Tao, T.: Signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59 (8) (2005), 1207–1223.
- [3] Takhar, D. *et al.*: A new compressive imaging camera architecture using optical-domain compression, in *Proc. IS&T/SPIE Computational Imaging IV*, January 2006, 1–10.
- [4] Duarte, M.F. *et al.*: Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.*, 25 (2) (2008), 83–91.
- [5] Zomet, A.; Nayar, S.K.: Lensless imaging with a controllable aperture, in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2006, 339–346.
- [6] Chan, W.L.; Charan, K.; Takhar, D.; Kelly, K.F.; Baraniuk, R.G.; Mittleman, D.M.: A single-pixel terahertz imaging system based on compressed sensing. *Appl. Phys. Lett.*, 93 (12) (2008), 121105-1–121105-3.
- [7] Shrekenhamer, D.; Watts, C.M.; Padilla, W.J.: Terahertz single pixel imaging with an optically controlled dynamic spatial light modulator. *Opt. Exp.*, 21 (10) (2013), 12507.
- [8] Babacan, S.D. *et al.*: Compressive passive millimeter-wave imaging, in *2011 18th IEEE Int. Conf. on Image Processing (ICIP)*, September 2011, 2705–2708.
- [9] Jiang, H.; Deng, W.; Shen, Z.: Surveillance video processing using compressive sensing. *Inverse Probl. Imag.*, 6 (2) (2012), 201–214.
- [10] Kirmani, A.; Colaco, A.; Wong, F.N.C.; Goyal, V.K.: Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Opt. Express*, 19 (22) (2011), 21485–21507.
- [11] Howland, G.A.; Dixon, P.B.; Howell, J.C.: Photon-counting compressive sensing laser radar for 3D imaging. *Appl. Opt.*, 50 (31) (2011), 5917–5920.
- [12] Colaco, A.; Kirmani, A.; Howland, G.A.; Howell, J.C.; Goyal, V.K.: Compressive depth map acquisition using a single photon-counting detector: parametric signal processing meets sparsity, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2012, 96–102.
- [13] Caroli, E.; Stephen, J.B.; Di Cocco, G.; Natalucci, L.; Spizzichino, A.: Coded aperture imaging in X- and Gamma-ray astronomy. *Space Sci. Rev.*, 45 (1987), 349–403.
- [14] Jean in't Zand.: Coded aperture camera imaging concept, 1992, http://astrophysics.gsfc.nasa.gov/cai/coded_intr.html, retrieved 1 October 2013.
- [15] Li, C.; Jiang, H.; Wilford, P.; Scheutzw, M.: A new compressive video sensing framework for mobile broadcast. *IEEE Trans. Broadcast.*, 59 (11) (2013), 197–205.

Hong Jiang received his B.S. from Southwestern Jiaotong University, M. Math from University of Waterloo, and Ph.D. from the University of Alberta. Hong Jiang is a researcher and project leader with Alcatel-Lucent Bell Labs, Murray Hill, New Jersey. His research interests include signal processing, digital communications, and image and video compression. He invented key algorithms for VSB demodulation and HDTV video processing in the ATSC system, which won a Technology and Engineering Emmy Award. He pioneered hierarchical modulation for satellite communication that resulted in commercialization of video transmission. He has published more than 100 papers and patents in digital communications and video processing.

Gang Huang received his B.S.E.E. (with highest distinction), and the M.S.E.E. degrees from the University of Iowa, Iowa City, Iowa. Mr. Huang joined Alcatel-Lucent Bell Labs (formerly AT&T Bell Labs) in 1987. He is currently with the Video Analysis and Coding Research Department of Bell Labs. His current interests include Compressive sensing and its

application in image and video acquisition and compression. He holds more than 10 US patents in areas ranging from data networking, data transmission, signal processing, and image processing.

Paul A. Wilford received his B.S. and M.S. in Electrical Engineering Cornell University in 1978 and 1979. His research focus was communication theory and predictive coding. Mr. Wilford is a Bell Labs Fellow, and Senior Director of the Video Analysis and Coding Research Department of Bell Labs. He has made extensive contributions in the development of digital video processing and multi-media transport technology. He was a key leader in the development of Lucent's first HDTV broadcast encoder and decoder. Under his leadership, Bell Laboratories then developed the world's first MPEG2 encoder. He has made fundamental contributions in the high speed optical transmission area. Currently he is leading a department working on Next Generation video transport systems, hybrid satellite-terrestrial networks, and high-speed mobility networks.