Multi-dimensional Compressive Imaging


* Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, 06269 USA; b Lockheed Martin Missiles and Fire Control, Orlando, Florida 32819, USA; c Department of Electro-Optics Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel; d Department of Information and Physical Sciences, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan; e Institute of New Imaging Technologies, Universidad Jaume I Campus del Riu Sec s/n, 12071 Castellón de la Plana, Spain; f University of Valencia, Department of Optics, Valencia, Spain

Email: Email: Bahram.Javidi@uconn.edu

ABSTRACT

In this keynote address paper, we present an overview of our previously published work on using compressive sensing in multi-dimensional imaging. We shall examine a variety of multi dimensional imaging approaches and applications, including 3D multi modal imaging integrated with polarimetric and multi spectral imaging, integral imaging and digital holography. This Keynote Address paper is an overview of our previously reported work on 3D imaging with compressive sensing.

Keywords: Compressive sensing, integral imaging, spectral imaging

1. PRINCIPLES AND PERFORMANCE ANALYSIS OF INTEGRAL IMAGING

Stereoscopic and auto-stereoscopic 3D display technologies are based on the use of special glasses [1-4] or monitors [5-8] that in both cases send slightly different images to the left and to the right eyes to produce binocular disparity. All of these techniques have an essential problem, i.e., the conflict between eye accommodation and convergence of the visual axes. As a result, visual fatigue and sometimes strong feelings of discomfort occur [9]-[10].

A very interesting alternative proposal comes from a technique called Integral Photography, conceived by Gabriel Lippmann in 1908 [11]. It is based on the idea that it is possible to record a 3D image of a scene taking many pictures of it from different perspectives. This can be done on a macroscopic scale using an array of cameras or, in a smaller scale, inserting a microlens array (MLA) in front of the optical sensor (see Figure 1).

![Figure 1.- Capture procedure in integral imaging.](image-url)
The array of microlenses permits the acquisition of the 3D scene from many different perspectives. The individual images are usually named as elemental images. To avoid the overlapping between neighbor elemental images it is necessary the insertion of barriers between them. There is only one plane of the 3D scene, the conjugate plane, that produces sharp images onto the sensor. However, this blurring is usually negligible when compared with the size of the sensor pixels. Thus, in what follows we will consider that all parts of the 3D scene are captured sharply. To acquire enough 3D information it is necessary that any part of the scene be captured by many elemental images.

This 3D information can be processed in many different ways to be used in many interesting applications. Now we concentrate in the original application reported by Lippmann. This idea is to project the integral image onto a 2D display placed in front of a microlens array. The display and the microlenses used in the display should be similar to the ones used in the acquisition, or could be scaled proportionally. As shown in Figure 2, the different perspectives are integrated into a 3D image. Contrary to what happens in the auto-stereoscopic screens, now there is a real reconstruction of the light structure produced by the original 3D scene. Given an object point, in the reconstruction the ray bundles produced by the pixels of the display intersect at the same position of the original object. Then, what the observer receives is a diverging ray beam totally equivalent to that produced by a real point source. In this case, there is concordance between the accommodation effort and the convergence. Naturally, although this concept is already centenary, it is only very recent the ability of technology to produce suitable microlens arrays, and systems to acquire, display, process and transmit this information [12].

![Figure 2.- Display procedure in integral imaging.](image)

Over the last decade there has been a very active development of this technology, seeking to improve its performance resolution, visual angle, continuity of perspectives, or, among other applications, the ability to reconstruct 3D scenes, 3D shape recognition, or imaging under very low levels of illumination. In references [13]-[29] we summarize some of the most important contributions.

The original Lippmann concept is based on the acquisition of many perspectives of a 3D scene by means of a multilens recording of a 3D scene. The proper selection of the acquisition parameters strongly depends on the application. For example, when the aim is the recording of elemental images intended for being displayed in an integral imaging monitor, one has to take into account that microlens pitch is the display resolution unit (DRU) in integral imaging displays [30]. Thus, for this kind of application a large number of elemental images with moderate number of pixels are required.

It may happen that the 3D scene is far from the camera, so that the angular extension of the array of lenses, as seen from the center of the scene, is very small. In this case, a camera lens, also named as depth-control lens [31],[32], is necessary to image the reference plane of the far 3D scene onto the MLA. In that case, some parts of the 3D scene are imaged in front of the MLA and other parts behind the MLA. Since this capture modality, shown in Fig. 3, is essentially different from the one described above, it receives a specific name: the plenoptic camera [33].
Using the camera lens has the effect of transposing the resolution constraints [34]. Thus, in the plenoptic camera the MLA pitch determines the spatial resolution of reconstructed sections of the 3D scene. The angular resolution, or segmentation capacity of the 3D reconstruction, is restricted by the number of pixels per elemental image. From the captured elemental images one can calculate the so-called sub-images, or view-images, [35] by extracting and composing the pixels at the same local position in every elemental image.

This direct pickup procedure is very useful because it allows the acquisition of the elemental images by use of only one sensor and after only one snapshot. The parallax obtained with it is determined by the angle subtended by the camera lens as seen from the center of the scene. The plenoptic images captured by this procedure can be very useful for the depth reconstruction of far scenes. Also, since plenoptic cameras record scenes that are in the close neighborhood of the MLA, the acquired elemental images are ready for direct display in an integral imaging monitor.

Lenslet-based integral imaging systems suffer, however, from a limitation in the spatial resolution due to diffraction effects. (because lenslet-based integral imaging systems have a small numerical aperture). Three parameters come into play: the camera pixel size, the lenslet point spread function, and the lenslet depth of focus [36]. However, integral imaging can also be performed using a single 2D imaging sensor, scanning the aperture and capturing a discrete number of images over a large area. This approach is known as synthetic aperture integral imaging and overcomes some of the limitations of traditional lenslet-based integral imaging systems [37]. Other data capture approaches may be used as well. For example, see [38].

2. MULTIDIMENSIONAL COMpressive IMAGING

Integral imaging is a good platform for compressive sensing (CS) to acquire multi-dimensional optical information, including depth, spectrum, polarization, etc., of a scene. Since an integral image presents a high degree of redundancy, this redundancy can be used to acquire additional information. For example, the modulation of the optical signal using different types of pupils and sparse random sampling with a filter array on the image sensor have been proposed and demonstrated to acquire this information [39-41]. The coding schemes are explained in the following where spectral imaging is used as an example.

The first coding scheme for spectral integral imaging uses basically a spectral dispersion mechanism [39]. Different dispersers are located in the each element as shown in Fig. 4(a). The second coding scheme uses spectral filters to multiply the spectral datacube by different weight distributions in each element as shown in Fig. 4(b) [40]. The datacube is filtered with different pass/stop-bands. In the figure, multiple filters are located in front of each pupil to modulate the spectral bands. The third coding scheme is a pixel-wise color filter array on the image sensor. In this case, the datacube is sparsely and randomly sampled, as shown in Fig. 4(c). The sampling patterns in each element are different to reduce their redundancy [41].

Based on the third coding scheme, a multi-dimensional imaging system by using an integral imaging optics and an image sensor with randomly arranged pixel-wise filtering elements has been proposed in [41]. The imaging process of the entire system can be described as

\[ g = Hf \]  (1)
where $\mathbf{g}$ is the vector of the acquired data, $\mathbf{H}$ is the system model matrix that describes the imaging process, and $\mathbf{f}$ represents the vector of the object data. Because of the assumption that the dimensions of $\mathbf{g}$ are much smaller than the dimensions of $\mathbf{f}$, the system (Eq. (1)) is ill-posed. Because the system model matrix $\mathbf{H}$ has multiple nonzero elements, we can solve the inverse problem (Eq. (1)) by using a compressive sensing algorithm called Two-step iterative shrinkage/thresholding (TwIST) [42] as follows:

$$
\hat{\mathbf{f}} = \text{arg min}_{\mathbf{f}} \| \mathbf{g} - \mathbf{Hf} \|_2^2 + \tau R(\mathbf{f})
$$

where $\| . \|_2$ represents the $l_2$ norm, $\tau$ is a regularization parameter, and $R(\cdot)$ is a regularizer. Total variation [43] can be chosen as the regularizer.

![Figure 4. Compressive spectral integral imaging. Pupil engineering with (a) shearing and (b) weighting. (c) Sparse sampling with a filter array.](image)

The experimental result for a depth-dependent multispectral object with the coding scheme of Figure 4(c) is shown in Figure 5 [41]. In the demonstration, the conventional back-projection and a TwIST algorithm were used. In the TwIST reconstruction, the four-dimensional object including two depths and three spectra was successfully reconstructed.

![Figure 5. Experimental reconstruction with (a) the conventional back-projection and (b) the TwIST.](image)

The experimental results by applying TwIST to a multi-dimensional polarimetric integral imaging system [44] have been also reported [41]. The reconstructed images with the back-projection and the TwIST methods are shown in Figure 6, respectively. From the experimental results, we can say that the TwIST algorithm suppressed the defocused objects compared to the conventional back-projection algorithm. Furthermore, TwIST enhanced the contrast and lateral resolution of the reconstructed object.
3. COMPRESSIVE DIGITAL HOLOGRAPHIC SENSING

Compressive sensing has been recently and successfully combined with digital holography, yielding new applications and solutions to classical holographic imaging problems. A review of such applications can be found in [45].

Digital holography applications using compressive sensing paradigm tools can be divided coarsely into three scenarios, corresponding to three subsampling schemes. The first one is random subsampling of the Fresnel field. Suppose that we wish to capture only $M$ spatial measurements of the Fresnel field. For instance, in order to design a system where our detector budget is constrained to $M$ detectors and the signal we need to reconstruct consists of $N$ samples, but it only has $S$ meaningful entries. It is shown in [46] that a signal can be accurately reconstructed if the number of compressive samples, $M$, obeys:

$$M \geq \begin{cases} \frac{CN^2}{N} \frac{S}{\log N} & \text{for } z \leq \sqrt{N} \frac{\Delta^2}{\lambda} \\ CS \log N & \text{for } z > \sqrt{N} \frac{\Delta^2}{\lambda} \end{cases}$$

where $N_p = \sqrt{N \Delta / (\lambda z)}$ denotes the recording device Fresnel number, $\Delta$ is the pixel size, $\lambda$ is the wavelength, $z$ is the distance between the recording device and the object planes, $N$ is the number of pixels and $C$ is a small numerical constant. This first inequality in Eq. (3), referring to the near field numerical approximation ($z \leq \sqrt{N} \frac{\Delta^2}{\lambda}$) [47], expresses the intuitively expected dependence of the number of samples on the physical properties of the sensor, on the wavelength of the light and on the reconstruction distance. As for the far field ($z > \sqrt{N} \frac{\Delta^2}{\lambda}$) numerical approximation in the second inequality in Eq.(3), the number of compressive samples is constant for every $z$ and is the same as the results obtained from random subsampling of Fourier transform sensing mechanism [48].

The second subsampling scheme corresponds to deterministic subsampling, which may be the most suitable in many physical scenarios where the object’s wavefront is partially distorted by, for example, an occluding media, finite aperture of a lens or turbid media. After passing through this media, the wavefront is distorted and details are considered to be lost, when employing classical arguments. It was shown in [49] that the forward sensing model of this problem can be described as a deterministic subsampling of the object’s Fresnel wave propagation. Therefore, object reconstruction guarantees may be formulated using tools developed in CS theory. In [49] authors showed that the number of sparse features, $S$, which can be accurately reconstructed, is given by:
where $p(n)$ is a function that describes the distorting media, $F$ denotes the discrete Fourier transform and $(\cdot)^*$ denotes the complex conjugate. Using the suggested framework it was demonstrated in a real experiment [49] that an object with 59% of its wavefront truncated by an occluding plane can be reconstructed with dramatically improved accuracy, compared with standard numerical Fresnel back propagation. Equation (4) can also be used to derive super-resolution guarantees, for example when the wavefront goes through a finite aperture [50].

While in the previous two subsampling scenarios we have discussed the accurate reconstruction of 2D objects, from their 2D recorded wavefield, the third subsampling scenario considers accurate reconstruction of a 3D object from its 2D recorded wavefield, which is one of the main benefits of holography. Suppose that we have a 3D volume consisting of $N_x \times N_y \times N_z$ voxels, with each voxel size being $\Delta x \times \Delta y \times \Delta z$, is being mapped to a 2D detector with $N_x \times N_y$ pixels, where $N_z$ is the number of object planes. Assuming the Born approximation [51], this mapping is simply described as a superposition of each object plane convolved with its corresponding Fresnel kernel (which depends on the distance from that plane to the detector plane). It was demonstrated in numerous works that it is indeed possible to extract the 3D information from a 2D measurement [45, 51]. An example is shown in Figure 7. The original 3D object, which is composed of 4 planes separated $\Delta z$ apart, is shown in Figure 7(a). If we record the wavefield of the object and attempt to reconstruct it using standard numerical back propagation (refocusing), it will yield the result shown in Figure 7(b), where each reconstructed plane is distorted, due to object points which do not belong to the same object plane. The result shown in Figure 7(c) illustrates the compressive reconstruction from the same recorded object wavefield.

$$S \leq \frac{1}{2} \left(1 + \frac{|F(p[n])|^2}{\max_{m \neq 0} \sum_n |F(p[n])F^*(p[n-m])|} \right)$$  \hspace{1cm} (4)$$

Recently, it was shown that for this sampling scenario, the number of sparse signal elements which can be accurately reconstructed is given by the following equation [52]:

$$S \leq 0.5 \left(1 + \lambda \Delta z / \Delta^2 \right).$$  \hspace{1cm} (5)$$
This result formulates the reconstruction guarantees for performing object tomography of a 3D object from its 2D recorded hologram, and in some cases predicts object reconstruction accuracy beyond standard diffraction limits [52].

4. MULTISPECTRAL INTEGRAL IMAGING

Integration of spectral information into 3D systems has applications in many fields, including detection of illnesses [53], and remote sensing, among others. A multispectral integral imaging system is proposed in [54]. This system consists of a Marlin F080B model camera, whose CCD sensor size is 4.80×3.62mm (and the image size is 1032×778 pixels) and a Liquid Crystal Tunable Filter (LCTF) which can acquire a maximum of 33 bands in the [400,720]nm spectral range, with a spectral resolution of 10nm. A zoom lens is screwed to the rear part of the filter, and a makro system is used between the filter and the camera. A fiber optic illumination system (halogen lamp) with a diffuser is also used. See Figure 8 for details.

![Multispectral integral imaging acquisition experimental set-up.](image)

![3D profile of the 3 dice scene for λ=550nm](image)

Moving the scene to be acquired into a series of positions of the grid and moving the camera in an array of positions have an equivalent effect. Due to the weight of the optical acquisition system and the specifications of the motors, the
scene was finally moved in a regular 11×11 grid. Only 7 wavelengths spanning the [480, 680] nm wavelength interval were considered, due to the limitation of the integration times. Figure 9 shows the corresponding 3D profile of the dice scene that can be seen in Figure 8, for λ = 550nm. This profile is obtained by applying a depth variance minimization method originally proposed in [55]. The z axis represents the distance (in mm) from the acquisition system to the objects that define the scene. The x and y axes represent the lateral pixels of the scene. We can infer from the 3D profile that each plane corresponding to closest face of each dice is at a different and well defined distance.

5. CONCLUSIONS

In this paper we have presented an overview of some of our previously reported research that involve the application of compressive sensing in multi-dimensional imaging, in particular, 3D multi modal imaging integrated with polarimetric and multi spectral imaging, integral imaging, and digital holography. The results are also presented to highlight the potential applications and future research lines that could be derived from them.

REFERENCES


