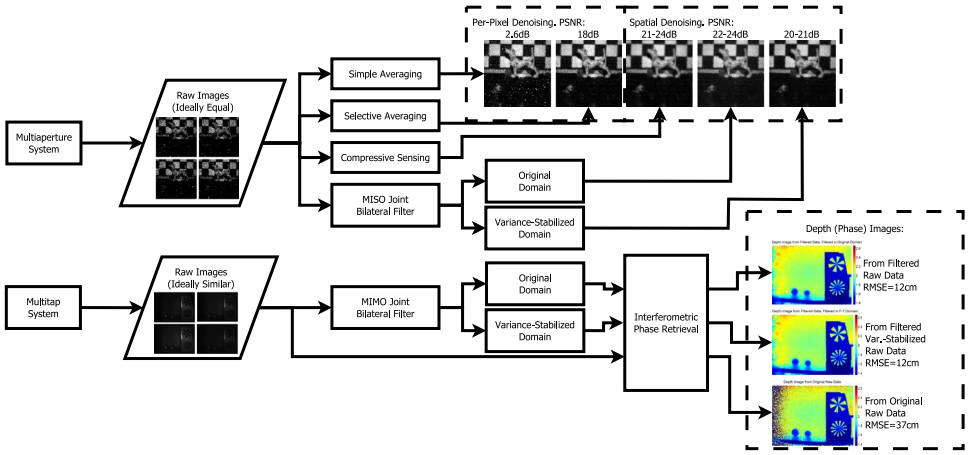
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

# Low-Light Image Enhancement for Multiaperture and Multitap Systems

Volume 8, Number 2, April 2016

**Miguel Heredia Conde**
**Bo Zhang**
**Keiichiro Kagawa**
**Otmar Loffeld, Senior Member, IEEE**




DOI: 10.1109/JPHOT.2016.2528122
1943-0655 

# Low-Light Image Enhancement for Multiaperture and Multitap Systems

**Miguel Heredia Conde,[1] Bo Zhang,[2] Keiichiro Kagawa,[2] and Otmar Loffeld,[1] *Senior Member, IEEE***

[1]Center for Sensorsystems (ZESS), University of Siegen, 57076 Siegen, Germany
[2]Imaging Devices Laboratory, Research Institute of Electronics, Hamamatsu 432-8011, Japan

*DOI: 10.1109/JPHOT.2016.2528122*


Manuscript received November 2, 2015; revised February 3, 2016; accepted February 6, 2016. Date of publication February 11, 2016; date of current version March 1, 2016. This work was supported by the German Research Foundation (DFG) as part of the research training group GRK 1564 "Imaging New Modalities." Corresponding author: M. Heredia Conde (e-mail: heredia@zess.uni-siegen.de).

**Abstract:** Intense Poisson noise drastically degrades image quality when only a few or when a single photon hits each pixel. Multiaperture systems are able to provide multiple images of the same scene, which are acquired simultaneously. After registration and cropping, the partial scene information contained in each aperture image should be the same, while the noise will be different in each one. A similar case arises in multitap systems, which are widely used in Time-of-Flight imaging (ToF), where several integration channels per pixel exist and where several sequential acquisitions are needed to generate a depth image. In this case, raw images might be different from each other, but still, since they are images of the same scene, information redundancy can be exploited to filter out the noise. In this work, we propose two different ways of joint processing of low-light multiaperture images. One of them is an extension of bilateral filtering to the multiaperture case, while the other relies on the compressive sensing theory and aims to recover a noiseless image from fewer measurements than the total number of pixels in the original noisy images. Experimental results show that both methods exhibit very close performance, which is much higher than those of previous methods. Additionally, we show that bilateral filtering can also be applied to the raw images of multitap ToF systems, leading to a significant error reduction in the final depth image.

**Index Terms:** Low-light, compressed sensing, Time-of-Flight (ToF), multiaperture, multitap.

## 1. Introduction

Low-light conditions are understood to be a level of light for which a certain sensing device cannot resolve the signal because it is within its noise floor. The sensing device can be, for instance, a human eye or a camera. In general, the energy captured by a sensing element (e.g., cones and rods in a human eye or pixels in a camera) depends on the power density at its surface, the exposure time, the sensitive area of the element and its own sensing efficiency. Increasing the exposure time is always possible in a camera, but at the cost of loosing time resolution, which will result in image blur in non-static scenes. The power density can be increased increasing the aperture of the lens in a camera or dilating the pupil in a human eye. Nevertheless, a physical limit exists in both cases and the energy stored by the sensing element is then limited by scene illumination.

Imaging in low-light conditions is an appealing topic that has attracted much attention from the very beginning of digital imaging. Even before the advent of digital cameras, a considerable effort had been made to provide the ability to *see* in the darkness. Early devices to this end were bulky image intensifier tubes, whose development started in the 1930s and were able to generate an image on a phosphor screen. Pushed by the military, research on image intensifiers has provided an uninterrupted improvement of their sensitivity, while the image distortion, noise and size of the tube kept shrinking. Nowadays, image intensifiers with standard mounts can be attached to a conventional camera as an intermediate lens.

Despite image intensifiers offer nowadays an acceptable size and low image distortion, they are active systems, requiring high voltages and with a non-negligible power consumption. Charge coupled device (CCD) cameras offer outstanding sensitivity and low noise and can be used for imaging in low-light conditions. The so-called intensified-CCD sensors combine an image intensifier with a CCD array, while the recent *electron-bombarded* CCDs offer a compact solution, in which the phosphor screen has been eliminated and the accelerated electrons impact directly on the backside of the CCD. Noise can also be significantly reduced by cooling the camera. These technologies are widely used in medical devices and especial applications, eventually allowing single-photon-imaging but are not easily implementable in low-cost imaging devices.

An appealing alternative to achieve good quality images in low-light conditions is the use of multiaperture systems. Several multiaperture images can be combined to generate an image with improved SNR, which retains the signal contained in the raw images, while discarding the noise. Multiaperture cameras can be used to capture fast phenomena in low-light conditions without external intensifiers, since they can reduce the noise without increasing the exposure time.

Multitap image sensors, which are often used in Time-of-Flight (ToF) imaging, are close to multiaperture systems in that they generate several images of the same scene per acquisition. The difference lies in the fact that the multitap images are, in general, not expected to be equal to each other, but similar, while the noise will differ from one image to another. As well as in the multiaperture case, the similarity of the raw image can be exploited to filter out the noise. Especially in ToF imaging, operation in low-light conditions is highly desirable. ToF systems are active systems, i.e., they are equipped with an illumination unit, so that they receive the reflection of the light they emit. The exposure time is typically limited by motion blur and frame rate requirements. Therefore, in real scenes with large depth ranges, provided that the optical power of the illumination cannot be arbitrarily increased, the ability of operating in low-light conditions is crucial to provide an acceptable depth estimation in areas of the scene from which little light returns to the camera.

Both in the multiaperture and multitap cases, the way the similarity between raw images is exploited determines how well the noise can be removed and the cost in terms of information loss. In this paper we present two methods for combining the images of a multiaperture camera. The first one is based on bilateral filtering, while the second exploits the sparsity of real images in gradient domain to recover a denoised image from few noisy measurements in a compressive sensing (CS) framework. We show that the proposed methods achieve better performance than simpler state-of-the-art approaches. A joint bilateral filter is considered as a simple alternative to existing CS-based approaches for denoising of multitap images. We show that the method allows obtaining depth maps of better quality from real raw data of a ToF camera, acquired in low-light conditions.

## 2. Hardware Description and State-of-the-Art Multiaperture Denoising

### 2.1. Multiaperture Imaging

The schematic drawing of a multiaperture imaging system is shown in Fig. 1. It consists of an array of image sensors and lenses. One lens and one sensor constitute an aperture, which can

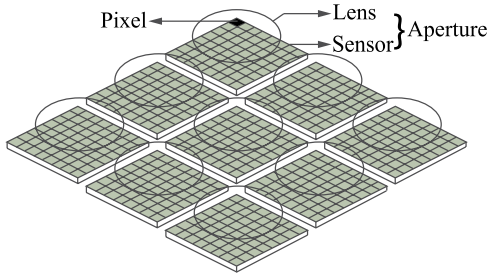



Fig. 1. Structure of a multiaperture imaging system.

be considered as a conventional camera. In a single shot, multiple images are acquired simultaneously. The image size is typically the same for all apertures.

Camera arrays and multiaperture imaging systems have been used to boost the capabilities of conventional imaging systems. Representative achievements are the synthesis of high dynamic range images [1], superresolution and refocus [2], ultra-high-speed imaging [3], and system miniaturization [4].

In this work, a multiaperture imaging system is used for noise reduction based on a plurality of images, acquired simultaneously. The images are combined into a single image using some noise-reduction method, e.g., selective averaging, to improve the image quality in low-light conditions. In the case of single-aperture imaging system under low-light conditions, large lens systems are needed to collect as much light as possible and correct aberrations, respectively. In the case of a multiaperture imaging system, however, it is possible to reduce the size and weight of lens, as well as to increase the imager sensitivity by means of a lens array. The synthetic $F$-number of multiaperture imaging system can be obtained as $F_s = F_0/\sqrt{M}$, where $F_0$ is the $F$-number of each elemental lens, and $M$ is the number of apertures.

### 2.2. Multitap Sensors and Time-of-Flight Imaging

Multitap imaging systems are those implementing pixels with two or more integration channels, often referred as *taps*. The electrons generated in the pixel by the incoming photons are integrated in one of the taps. The selection of the active tap is typically determined by a control signal and is, therefore, time-dependent. This means that, during the integration time, the photogenerated electrons contribute to one pixel tap or another, depending on when they were generated. Consequently, the main goal of these pixels is to demodulate an intensity modulated radiation. The first multitap pixels were presented in [5] and [6]. The idea of a lock-in structure was immediately adopted to generate the first 2-D arrays of demodulating pixels for ToF imaging [7]. Examples of pixel configurations are 2-tap and 4-tap, the former being the option adopted in most commercial systems. The first ToF imaging sensor implementing these *smart* pixels was called the photonic mixer device (PMD) and became the reference technology for phase-shift ToF imaging [8]. A survey of lock-in ToF cameras can be found in [9]. The advent of first Kinect sensor, with higher native lateral resolution than PMD sensors motivated research on the relative performance of multitap ToF systems, with respect to the novel light coding technology implemented in the Kinect sensor. A thorough comparison of both depth imaging technologies [10] revealed that, at the time the Kinect sensor was released, PMD cameras offered better effective angular resolution, despite the larger number of pixels of the Kinect. This might be one of the reasons that motivated the migration from structured light to ToF technology in the new generation of the Kinect sensor, commercialized as Xbox One sensor. The new sensor features 2-tap pixels, i.e., two integration channels per pixel, similarly to PMD chips. An improved pixel design allows for a high fill factor and, together with fast and numerous ADCs, make possible to integrate a large number of pixels in the sensor [11]. In

consequence, nowadays, multitap systems constitute the state-of-the-art technology for depth sensing.

Phase-shift-based ToF sensors estimate the depth from the phase shift between the emitted modulation waveform and the corresponding reflection, as received at the pixel surface. The illumination signals are periodic signals, with base frequencies in the megahertz range. The multitap pixels act as correlators, since they compute the cross-correlation between the received light signal and the reference signal which regulates the integration process in each tap. For the sake of generality and provided that the Xbox One sensor is a proprietary hardware that does not allow for external adjustment or modifications, we adopt the PMD technology as reference technology for multitap ToF imaging. Nevertheless, the method we present in this paper for multitap low-light image enhancement is fully applicable to the raw data of the Xbox One sensor. In the case of PMD, the reference signal controlling the integration process is a shifted version of the so-called illumination control signal (ICS), which is, in turn, used to modulate the illumination. Consequently, after integrating along many periods, we obtain a sample of the autocorrelation function, at a certain phase, which depends from the phase shift induced by the depth and the phase shift between ICS and reference signal, which is known. Since the phase shift induced by the depth is, for a static scene, constant, varying the relative phase shift between illumination and integration control signals allows gathering samples of the autocorrelation function at different phases.

A common hypothesis that vastly simplifies the depth estimation is to suppose that the illumination signal is sinusoidal. Under such conditions, it can be assumed that the correlation is between sinusoidal signals, since high-frequency harmonics contained in the close-to-square reference signal lead to a null contribution when correlating with a pure sine. Despite it is arguable that a perfect sinusoidal illumination can be achieved with the conventional illumination systems of commercial ToF cameras, such hypothesis reduces the degrees of freedom of the problem to three [8], namely, a constant offset, given by the DC component of the light, the amplitude of the modulated signal and the phase shift induced on it by the depth. Only three correlation measurements suffice to determine the phase shift, but typically, four equidistant samples of the autocorrelation function are acquired, using four relative phase displacements of the integration signal with respect to the modulation signal $(\theta \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\})$. Then the depth can be estimated using the so-called *four phases algorithm*, given by

$$d = \frac{c}{4\pi f_{\text{mod}}} \arctan\left(\frac{D(270^\circ) - D(90^\circ)}{D(180^\circ) - D(0^\circ)}\right) \tag{1}$$

where $c$ is the speed of light in vacuum, $f_{\text{mod}}$ is the frequency of the modulation signal, and $D(\theta)$ is the difference between the levels of the pixel channels at the end of the integration time, for a certain phase displacement of the integration signal with respect to the modulation signal $\theta$. That is, using this method and 2-tap pixels, four acquisitions are required to generate a depth image. Using 4-tap pixels would reduce the number of acquisitions, at the cost of pixel complexity (and size) and lower fill factor. Note the strong non-linearity of the depth estimation, due to the arctan function. This is one of the main motivations of dealing with multitap systems in this work, since, under low-light conditions, the four phases algorithm might act as a non-linear noise amplifier. In other words, noisy raw images, where the scene is still distinguishable, might lead to completely wrong depth measurements and, in consequence, useless depth images. In this work we show that the redundancy among PMD raw images can be exploited to filter out noise in an intelligent way, at no cost in terms of power budget. Such noise reduction allows, in practice, reducing the exposure times while achieving equivalent depth errors.

### 2.3. Selective Averaging

In multiaperture imaging systems, all the images captured at the same time appear very similar. However, noise caused by the readout circuits of the sensor and dark current, as well as photon shot noise, differs from one pixel to another. Although the noise level can be reduced by

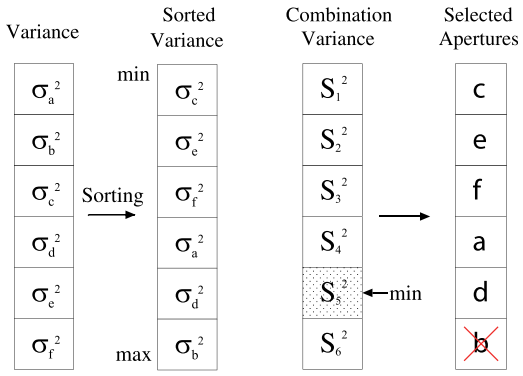



Fig. 2. Procedure of selective averaging.

averaging these noisy images, large random noise such as random telegraph signal (RTS) noise [12], photon shot noise and dark current shot noise cannot be eliminated. To remove the large random noise, the selective averaging method was proposed [13].

Unlike a simple averaging, the apertures which have large temporal noise are automatically excluded from the average by the selective averaging method. The aperture selection is operated one by one in the corresponding pixel of all apertures. Fig. 2 shows the procedure of aperture selection for one pixel. Firstly, the variance of the pixel value in darkness is calculated. The variances of the corresponding pixel in different apertures are sorted from minimum to maximum. After that, a combination variance is calculated by the following equation:

$$S_m^2 = \frac{1}{m^2}\sum_{i=1}^{m}\sigma_i^2 \tag{2}$$

where $m$ is the number of selected apertures, $\sigma_i^2$ is the sorted variance, i.e., $\sigma_i^2 < \sigma_{i+1}^2 \ \forall i < m$, and $S_m^2$ is the combination variance.

The apertures to be selected for a pixel are those for which $S_m^2$ is minimum. The value of each pixel in the selective averaging image is calculated by averaging the values of the corresponding pixel in the selected apertures.

## 3. Low-Light Image Enhancing

The method of selective averaging has shown to be an adequate way to cope with large levels of shot noise, present in images acquired in very low light conditions. The approach takes profit of the multiple images provided by a multiaperture camera to generate a single image with improved SNR. PSNR increments of 6.3 dB on real multiaperture images have been reported in [13]. Despite the promising results, the method works in a per-pixel manner and, therefore, does not exploit local correlations in the images. As a matter of fact, the amount of *information* contained in an image is not given by the number of pixels, but depends on the scene sensed. In other words, in most natural signals, an increasing volume of *data* might not suppose an equivalent increase in information terms, since correlations might exist in the data and patterns can be extracted. This is the basic idea behind compression. There is, consequently, room for improvement of the method in [13] if these considerations are taken into account.

A classic way of denoising taking profit of the local correlations present in natural images is filtering. Filtering in frequency domain is an easy way to remove noise of random nature, since it is mostly supported on the high frequencies. Filtering exclusively in frequency domain, e.g., convolving with a Gaussian kernel, comes at the cost of smoothing the image, eventually loosing structure and sharp edges. Bilateral filtering [14] solves the problem extending the filter to account not just for closeness in spatial domain but for similarity in the intensity domain as well.

In other words, the intensity of each pixel of the filtered image is a weighted sum of the intensity of neighboring pixels, where the weights depend on the distance in spatial domain and in intensity domain. This is just an integration of our *a priori* information on natural images, which, despite being piecewise smooth, might contain edges and texture. Bilateral filtering has been already used for image denoising in [15], where two images of a low-light scene, acquired with and without flash, are jointly filtered in order to transfer the detail of the flash image to the no-flash image, which correctly captures the colors but is corrupted by noise. Joint bilateral filtering can also be used for upsampling a low resolution image [16], if a high resolution modality is available.

A bilateral filter performs a convolution with an *adaptive* kernel is spatial domain, since the weights are no longer constant, but depend on the signal itself. This already brings the idea that, if we have deep knowledge on the nature of our signals, we might perform an optimal filtering. Adaptive methods exploiting piecewise constancy or piecewise polynomial or spline representation of natural images were first steps in this direction. Donoho *et al.* [17] showed that an appropriate wavelet shrinkage is equivalent to an optimal spatial adaptation. The underlying assumption behind soft-thresholding denoising [18] in wavelet domain is that the signal admits a sparse representation in some basis, e.g., the wavelet basis, i.e., that there exist a number of null or negligible coefficients in such representation. For a performance comparison of wavelet shrinkage estimators in the case of Poisson counts, see [19].

Intuitively, it is clear that a compressible signal should be easier to recover from noisy measurements than another, for which no basis is known where it can be sparsely represented. This idea has been formally studied by the recent theory of compressive sensing (CS) [20]–[22], which assures that a sparse or compressible signal can be exactly recovered from incomplete and noisy measurements if certain conditions are satisfied. In the context of our problem, this means that a number of measurements lower than the number of pixels might suffice to recover the denoised image if it admits a sparse or compressible representation. The measurements constrain the solution, while the recovery method has to be able to converge to the sparsest solution satisfying those constrains. Since noise does not admit a sparse representation in any structured dictionary, it is filtered out. CS imposes a linear sensing model, i.e., a *measurement* is a linear combination of pixel values, according to a certain *sensing kernel*. One of the main concerns of CS theory is the incoherence between such measurement kernels and the elements of the sparsity basis. Recoverability guarantees are subject to low coherence between the so-called *sensing matrix*, containing the sensing kernels, and the representation matrix, containing the sparsity basis. Consequently, random sensing kernels or periodic kernels can be adopted, since they are naturally incoherent with the wavelet basis. CS can be used as a general framework for image denoising [23], where the degrees of freedom left by a lower number of measurements than unknowns are typically constrained by the requirement of finding the sparsest solution in a given basis or dictionary.

In the following, we present both an extension of bilateral filtering to the case of multiaperture systems and a compressive sensing approach, which aims to recover the denoised image from few measurements, gathered using random binary kernels. Finally, we propose to apply bilateral filtering jointly to the raw images of multitap systems.

### 3.1. An Extension of Bilateral Filtering for Multiaperture Systems

A conventional bilateral filter [14] working on a single image takes into account the distance between each pixel and all the pixel contained in a certain neighborhood, both in spatial domain (closeness) and intensity domain (similarity), as shown in

$$\hat{I}(\vec{x}) = \frac{\sum\limits_{\vec{x}_i \in \Omega_{\vec{x}}} I(\vec{x}_i) w(\vec{x}_i, \vec{x})}{\sum\limits_{\vec{x}_i \in \Omega_{\vec{x}}} w(\vec{x}_i, \vec{x})}, \quad \text{where } w(\vec{x}_i, \vec{x}) = \exp\left( -\frac{\|\vec{x}_i - \vec{x}\|_2^2}{2\sigma_s^2} - \frac{\|I(\vec{x}_i) - I(\vec{x})\|_2^2}{2\sigma_i^2} \right) \tag{3}$$

where $\vec{x}$ are the coordinates of a pixel in the image, $\Omega_{\vec{x}}$ is a neighborhood of pixel $\vec{x}$, $I(\vec{x})$ is the intensity value of pixel $\vec{x}$, and $\hat{I}$ denotes the filtered image. The variables $\sigma_s$ and $\sigma_i$ are the smoothing parameters in spatial and intensity domain, respectively. The filter converges to a Gaussian filter when $\sigma_i \to \infty$.

Obviously, such a bilateral filter could be applied directly to the multiaperture images prior to a selective averaging or to the selective averaging result. Nevertheless, such approaches do not exploit the advantage of having multiple images of the same scene within the filter. To that end, the filter has to be extended to cope with several images while generating a single filtered output. We adopt the result of the selective averaging method, $I_{SA}$ as reference image, with respect to which distances in intensity domain are computed. As described in [13], $I_{SA}$ is computed as the average of the multiaperture images, considering, for each pixel, only a selected number of apertures. Suppose that we have $N$ apertures, $A_k, k \in [1, N]$. Let $I_k$ be the image gathered by the aperture $A_k$, after registration and cropping, so that any scene point corresponds to the same pixel in all multiaperture images. We denote $\Omega_k$ the set of all pixels for which the aperture $k$ was considered in the selective averaging. Then, (4), shown below, provides the proposed filter extension

$$\hat{I}(\vec{x}) = \frac{\sum_{k=1}^{N} \sum_{\vec{x}_i \in \Omega_{\vec{x}} \cap \Omega_k} I_k(\vec{x}_i) w_k(\vec{x}_i, \vec{x})}{\sum_{k=1}^{N} \sum_{\vec{x}_i \in \Omega_{\vec{x}} \cap \Omega_k} w_k(\vec{x}_i, \vec{x})}, \text{ where } w_k(\vec{x}_i, \vec{x}) = \exp\left( -\frac{\|\vec{x}_i - \vec{x}\|_2^2}{2\sigma_s^2} - \frac{\|I_k(\vec{x}_i) - I_{SA}(\vec{x})\|_2^2}{2\sigma_i^2} \right). \quad (4)$$

Note that the filter does not operate only in the spatial domain, but also along the apertures. By forcing the inner summation to consider only those neighborhood pixels contained in $\Omega_k$, we assure that the result of the filter cannot be worse than that of the selective averaging method, used as input, if $\sigma_s$ and $\sigma_i$ are appropriately selected.

### 3.2. Compressive Sensing for Multiaperture Systems

CS is a mathematical theory that provides conditions and limits for recovery of sparse or compressible signals from fewer measurements than those suggested by the Shannon sampling theorem. Natural images are known to be compressible in wavelet domain. This means that the denoising problem might be formulated as finding the sparsest representation in wavelet domain that agrees with the measurements. If the noise is independent from the signal and the SNR high enough, it will be automatically filtered out. Finding the sparsest solution to an underdetermined problem is known to be NP-hard [24]. Fortunately, it has been shown [25] that the corresponding linearly-constrained $l_0$ minimization can be substituted by an equivalent $l_1$ minimization if the sparsity is lower than a certain bound, related to the minimum ratio between the $l_1$ and $l_2$ norms of the vectors contained in the null space of the measurement matrix. The resulting linearly-constrained $l_1$ minimization can be efficiently solved as a linear program [26].

Let, $\vec{I_k} \in \mathbb{R}^n$, $k \in [1, N]$ be the images obtained from the $N$ apertures, being $n = n_{rows} \times n_{cols}$ the size of the vectorized raw images. Let $\vec{I} \in \mathbb{R}^n$ be the noiseless image we want to recover and $\vec{X} \in \mathbb{R}^n$ the corresponding sparse vector of coefficients in, e.g., wavelet domain. Let $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ be the sparsity basis or dictionary by columns, e.g., the wavelet basis, so that $\vec{I} = \mathbf{\Psi}\vec{X}$. Suppose that $N$ measurement vectors, $\vec{Y_k} \in \mathbb{R}^{m_k}$, are obtained from the $N$ raw images, using $N$ sensing matrices, $\mathbf{\Phi}_k \in \mathbb{R}^{m_k \times n}$, that is, $\vec{Y_k} = \mathbf{\Phi}_k \vec{I_k}$. The easiest way to merge the measurement vectors into a single vector is stacking them one after another and do the same with the sensing matrices, by rows. Other ways of combining the measurement vectors are also possible (e.g., averaging them to reduce uncorrelated noise) but require an equivalent

combination of the sensing matrices. If we adopt the stacking approach, we obtain the minimization problem

$$\hat{\vec{X}} = \underset{\vec{X}\in\mathbb{R}^n}{\operatorname{argmin}} \|\vec{X}\|_1 \text{ subject to } \vec{Y} = \boldsymbol{A}\vec{X},\ \vec{Y} = \begin{bmatrix} \vec{Y}_1 \\ \vec{Y}_2 \\ \vdots \\ \vec{Y}_N \end{bmatrix},\ \boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_2 \\ \vdots \\ \boldsymbol{\Phi}_N \end{bmatrix},\ \boldsymbol{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}. \tag{5}$$

Unfortunately, under conditions of very low light, the solution to (5) is not better than that obtained by simpler and faster methods, such as selective averaging. The reason is not an insufficient number of compressed measurements, but the combination of strong photon shot noise, dark current shot noise and RTS noise that corrupt the images. Strong punctual artifacts in the images, often much stronger than the underlying signal, are still present in the recovered image, by means of high-frequency wavelets localized at, or close to, that point. Strong sparsity demands lead to signal loss, instead of proper denoising. Increasing the number of measurements leads to better recovery of the noise. We have performed preliminary experiments using random sensing matrices of different nature, e.g., Gaussian, Bernoulli or a randomized Hadamard matrix, obtaining similar results for all cases. Obviously, gathering partial Fourier measurements provides an easy way to leave away undesired high frequencies and perform a filtering at sensing, but it also discards the possibility of recovering such frequencies in the signal.

For the low-light case, there is a restriction that is more meaningful than approximate sparsity in some dictionary: the total variation (TV). The total variation can be interpreted as a measure of the texture (or noise) in an image. Substituting the restriction on the $l_1$ minimization in (5) by a TV-minimization eliminates the dependency on a specific dictionary and requires sparsity in some more general *gradient domain*. The resulting optimization problem is rewritten in

$$\hat{\vec{I}} = \underset{\vec{I}\in\mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{n} \|\boldsymbol{G}_i\vec{I}\|_p + \frac{\mu}{2}\|\boldsymbol{\Phi}\vec{I} - \vec{Y}\|_2^2 \tag{6}$$

where $\boldsymbol{G}_i$ is the gradient operator centered at location $i$, $\mu$ a penalty parameter to be adjusted and $p = \{1, 2\}$. We observed no performance variation with the selection of $p$ and $p = 2$ was used for the experiments in this paper. Adopting this recovery framework, which is also close to the general image denoising framework of [23], allows achieving a good tradeoff between noise reduction and detail loss. It was observed that there is also not much difference in performance between averaging the vectors of measurements, i.e., $\vec{Y}_k = \boldsymbol{\Phi}_{\boldsymbol{k}}\vec{I}_k$, $\vec{Y} = (1/N)\sum_{i=1}^{N}\vec{Y}_k$, and measuring on the selective averaging result through a single sensing matrix, i.e., $\vec{Y} = \boldsymbol{\Phi}\vec{I_{SA}}$. On the one hand, performing measurements in each of the raw images separately has a pre-filtering effect, since uncorrelated noise will tend to be canceled out. On the other hand, the selective averaging result already provides better SNR and is a good starting point for the CS framework. We decide to combine both approaches, using the sets $\Omega_k, k \in [1, N]$ of all pixels for which the aperture $k$ was considered in the selective averaging to adapt each $\boldsymbol{\Phi}_{\boldsymbol{k}}$ to the signal to measure, $\vec{I}_k$. In short terms, we set to zero the columns corresponding to pixels for which the aperture was not considered in the selective averaging method, that is, $\vec{\Phi}_k^i = \vec{0}$ $\forall\ i \notin \Omega_k$, $\vec{\Phi}_k^i$ is the $i^{\text{th}}$ column of $\boldsymbol{\Phi}_{\boldsymbol{k}}$. This way we are performing a *selective compressed sensing*, still independently for each $\vec{I}_k$. Then, the measurements are averaged and the corresponding sensing matrix to be used in (6) is $\boldsymbol{\Phi} = (1/N)\sum_{i=1}^{N}\boldsymbol{\Phi}_{\boldsymbol{k}}$. The selection of the sensing matrix type does not affect much the recovery and both Fourier and random matrices delivered similar results. For simplicity, in the experiments presented in this paper, binary matrices with values $\phi_{i,j} \in \{-1, 1\}$ were used. They were obtained by randomizing a Hadamard matrix, after eventual cropping, and exhibit certain orthogonality by rows.

### 3.3. Bilateral Filtering for Multitap Systems

As discussed in Section 3.3, multitap systems are different from multiaperture systems, but closely related. Actually, since one of the main drawbacks of multitap systems is the low fill factor, a multiaperture sensor could be used to gather all the necessary images simultaneously, avoiding the need of many taps per pixel or many acquisitions. An example of hybrid hardware was presented in [3], which could be described as a 2-tap multiaperture system. Such system is able to integrate according to pseudorandom binary patterns during a very short exposure. Different codes are used for different apertures, and the result of the integration can be interpreted as the scalar product between the light signal and the binary code in time domain. CS can then be used to resolve a high frame rate image sequence from the few measurements, exploiting the fact that natural images exhibit a restricted total variation. Such hybrid architectures become an alternative to conventional multitap arrays when many measurements are required, like in ToF sensors (e.g., four acquisitions in a conventional PMD sensor or ten in the Xbox One sensor). Also, the hardware in [3] seems to be an adequate platform to implement the CS framework for PMD-based ToF imaging proposed in [27].

In this section we provide a bilateral filter for the case of multitap or hybrid multiaperture-multitap systems. In these cases, the raw images are expected to be different to each other, in general. Nevertheless, the underlying scene is—after eventual registration and cropping in hybrid systems—the same for all images and, consequently, they are expected to be highly correlated. The filter has to exploit this fact, while preserving each one of the images separately. In other words, the filter has to be a MIMO system, in contrast to the MISO approach given in (4). Multimodal bilateral filtering for ToF systems have been already studied in [28], where the filter operates with a vector of intensities per pixel, instead a single intensity value. In that work a ZESS MultiCam [29], [30] was used, providing registered depth and color images simultaneously. The depth was treated as another channel, together with the color channels. Provided that the depth image has a much lower resolution than the color modality, the bilateral filter was intended to transfer the high resolution from the color image to an upscaled version of the depth image. Despite the good results, the approach relies on the hypothesis that the depth modality is highly correlated with the color modality, i.e., texture in the color images corresponds to texture in depth. Although this is a valid assumption for many natural and man-made scenes, it does not always hold. A more meaningful alternative would be to filter before depth calculus, adding the raw images as additional intensity channels. We adopt the multichannel approach proposed in [28] for bilateral filtering of the raw images of multitap systems. Since we focus on the low-light scenario, we do not suppose having an additional high-resolution color image to include in the joint filtering. Note that color pixel arrays suffer from an additional loss of optical power due to the Bayer filter. Let $\vec{s}(\vec{x}) \in \mathbb{R}^N$ be the vector of raw intensities for the pixel indexed by $\vec{x}$, where $N$ is the number of integration channels (multiplied by the number of temporally-modulated apertures, in the hybrid case). The bilateral filter is formulated

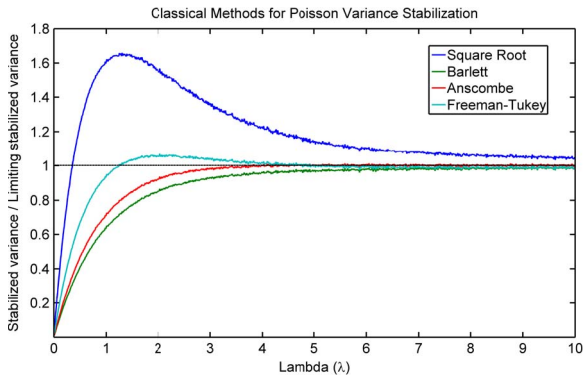
$$\hat{\vec{s}}(\vec{x}) = \frac{\sum_{\vec{x}_i \in \Omega_{\vec{x}}} \vec{s}(\vec{x}_i) w(\vec{x}_i, \vec{x})}{\sum_{\vec{x}_i \in \Omega_{\vec{x}}} w(\vec{x}_i, \vec{x})}$$

$$\text{where } w(\vec{x}_i, \vec{x}) = \exp\left( -\frac{\|\vec{x}_i - \vec{x}\|_2^2}{2\sigma_s^2} - \frac{(\vec{s}(\vec{x}_i) - \vec{s}(\vec{x}))^\top \Pi_{\vec{s}}^{-1} (\vec{s}(\vec{x}_i) - \vec{s}(\vec{x}))}{2} \right) \tag{7}$$

where $\Pi_{\vec{s}} \in \mathbb{R}^{N \times N}$ is the weighting matrix that accounts for the confidence of the intensity measurements. As in [28], we adopt a diagonal matrix, i.e., suppose statistical independence between channels. Since we work only with multitap raw data and for simplicity, we assign the same value, $\sigma_i^2$, to all the diagonal elements of $\Pi_{\vec{s}}$. This is an acceptable assumption in general, but note that its validity is signal-dependent. For example, in the case of PMD ToF, depths corresponding to phase shifts close to 0° and 180° might lead to very asymmetric charge distribution between

| Transformation | Formula |
|---|---|
| Square Root | $\sqrt{x}$ |
| Barlett | $\sqrt{x+\frac{1}{2}}$ |
| Anscombe | $\sqrt{x+\frac{3}{8}}$ |
| Freeman-Tukey | $\sqrt{x}+\sqrt{x+1}$ |

(a)




(b)

Fig. 3. Classic square-root-based variance stabilization transformations for Poisson data. The performance of the different methods in stabilizing the variance of synthetic Poisson data is given in (b). Freeman-Tukey transformation offers the best performance for data following Poisson distributions of very low $\lambda$ ($\lambda \leq 3$). (a) Variance stabilization formulas. (b) Performance comparison.

the two integration channels. In such cases, an independent choice of $\sigma^2_{i_k}, k \in [1, N]$ for each of the $N$ raw images is recommendable. In low-light imaging, this parameter can be tied to the average intensity level of the corresponding image. Note that, in most commercial ToF cameras, the low number of taps entails that several acquisitions are required to compute a depth image. In such cases, the full set of raw images can be jointly processed as in the hybrid case.

When there exist many integration channels or many acquisitions with different reference waveforms are required, CS arises as a natural framework to use the signal redundancy to discard uncorrelated noise. The application of CS to multitap systems is slightly more complex than the basic approach presented in Section 3.2 to deal with multiaperture data. The raw images differ from each other and, therefore, a joint processing in a Multiple Measurement Vector (MMV) framework becomes necessary. Additionally, if a *greedy* search is used as sparse recovery method, *a priori* knowledge on linear dependencies between raw data can be incorporated by restricting the rank of the residual matrix. A CS multichannel denoising approach for the case of a PMD sensor has already been presented in [31] and is out of the scope of this paper. The approach takes into account the peculiarities of the PMD technology to reduce the sensor data flow and achieve a dramatic noise reduction in the final depth image.

### 3.4. Denoising in an Appropriate Domain: Variance Stabilization

In low-light conditions, the noise contained in the image is dominated by shot noise, which follows a Poisson distribution. In this case, the variance of the noise is given by the signal itself, since in a Poisson distribution, the variance is equal to the expected value. In such conditions, we cannot assume that signal and noise are uncorrelated and the methods presented in Section 3 might not perform equally well along all image areas. In dark areas, the smoothing provided by such methods might wash away the signal, while being insufficient in bright areas, where the noise exhibits higher variance. There exist a number of variance-stabilizing transformations, which aim to stabilize the variance of data following a binomial or Poisson distribution. Some classic transformations based on the square root are the Barlett transformation [32], the Anscombe transformation [33], and the Freeman-Tukey transformation [34]. For completeness, we provide the corresponding formulas for stabilizing the variance of Poisson variables in Fig. 3(a), complemented with a plot showing the performance of each method for different values of $\lambda \in (0, 10]$ in Fig. 3(b). Recursive methods have been proposed to achieve optimal variance stabilization [35], taking classic stabilizing techniques, e.g., Anscombe and Freeman-Tukey transformations, as a starting point.

For expected values that are not too low, the square root transformation is the simplest option. For low expected values, the performance of the square root transformation degrades and an additional term within the square root is required to achieve good variance stabilization. Of this kind are the Barlett and the Anscombe transformations, performing the latter slightly better than the former. Unfortunately, in both cases the variance stabilization degrades rapidly when approaching to unit expected value. Note that a Poisson distribution with close-to-unit expected value matches our case of study, i.e., shot noise in very low-light imaging. In such cases, the Freeman-Tukey transformation provides notably better stabilization at the cost of a more complex expression that does not allow for a direct inversion.

#### 3.4.1. Freeman-Tukey Transformation

The Freeman-Tukey variance stabilization transformation for Poisson data has been given in Fig. 3(a), but this expression is not easily invertible and furthermore, it is a simplification of the more general double-arcsine formulation, for binomial distributions. Profiting from trigonometric identities, it has been shown that the Freeman-Tukey double-arcsine transformation [see (8) below] admits a closed-form inverse transformation [36], given by (9), shown below

$$t = \arcsin\sqrt{\frac{\lambda}{n+1}} + \arcsin\sqrt{\frac{\lambda+1}{n+1}} \tag{8}$$

$$\hat{p}(t) = \frac{1}{2}\left[1 - \operatorname{sgn}(\cos t)\sqrt{1 - \left[\sin t + \frac{1}{n}\left(\sin t - \frac{1}{\sin t}\right)\right]^2}\right] \tag{9}$$

where $p$ and $n$ are the parameters of the binomial distribution $B(n,p)$, namely, probability of success and number of independent experiments. Recall that the Poisson distribution is a limit case of the binomial distribution, when $p \to 0$ and $n \to \infty$. In the Poisson case, $P(\lambda)$, we are interested in the number of occurrences or successes, $\lambda = p\ n$. The direct Freeman-Tukey transformation for the Poisson case can be derived from (8) as $y = \lim_{n\to\infty}\sqrt{n+1}\ t$. Consequently, we can pursuit an inverse transformation $\hat{x}(y)$ for the Poisson case from (9) as $\hat{x}(y) = \lim_{\substack{(t=y/(\sqrt{n+1}))\\ n\to\infty}} n\ \hat{p}(t)$. This expression, after appropriate manipulation, leads to the inverse transformation we provide in (10), shown below, where $y$ is the transform, according to the Freeman-Tukey formula provided in Fig. 3(a), of the Poisson variable $x \sim P(\lambda)$. The complete proof takes profit from limit properties of trigonometric functions and is given in the Appendix.

$$\hat{x}(y) = \left(\frac{y - y^{-1}}{2}\right)^2 = \sinh^2 \ln y \tag{10}$$

According to our reprojection tests, the inverse transformation in (10) has shown to be exact up to machine precision. The experiments presented in Section 4 are always carried out in both domains, namely, the original intensity domain and the Freeman-Tukey domain. In the latter case, the results are evaluated in the original domain, after applying (10).

## 4. Experiments and Results

In this section, we describe the experiments we performed to evaluate the image enhancement achieved by the denoising methods presented in Section 3. The data is both real and synthetic, in order to allow for an exact quantitative evaluation of the improvement. The experiments are organized in two sections: Section 4.1 presents the experiments carried out on multiaperture data, while Section 4.2 provides an evaluation of the filtering on multitap data, in the interesting case of ToF imaging. In all cases, the experiments were carried out in the original image intensity domain and the variance-stabilized domain presented in Section 3.4.

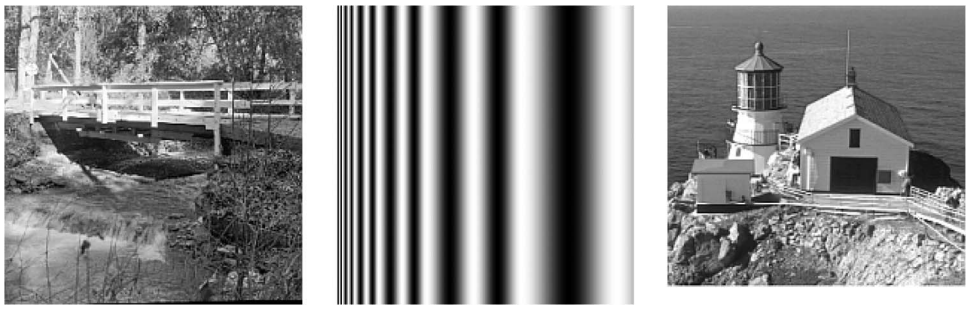


Fig. 4. Original images used to generate the low light multiaperture datasets. From left to right, the scenes are named “bridge,” “chirp,” and “lighthouse.”

### 4.1. Multiaperture Camera

Two methods were proposed in Section 3.1 and 3.2 to generate a noiseless image from a set of low light multiaperture images. The first method is an extension of a bilateral filter, while the second adopts a CS framework to recover the denoised image from measurements performed on the original multiaperture images. In this section, we provide an experimental evaluation of the performance of both methods using synthetic datasets and a real dataset, acquired with a multiaperture camera.

#### 4.1.1. Evaluation Using Synthetic Data

Three synthetic datasets are generated and assigned the names “bridge,” “chirp,” and “lighthouse,” which describe the content of their respective scenes. All the images are monochrome of size $200 \times 200$ pixels. The original noiseless images used to generate the datasets are given in Fig. 4 and are to be taken as reference images. The virtual multiaperture camera features $3 \times 3$ apertures, i.e., 9 multiaperture images per acquisition. For each dataset, three levels of light are considered, characterized by the maximum number of photons received by a pixel: $n_{e^-}^{\max} = \{1, 3, 9\}$. That is, in absence of any disturbance, each pixel would store a number of electrons $n_{e^-} \in [0, n_{e^-}^{\max}]$. In the dataset images this range is actually extended in both directions due to the addition of sensor noise and photon shot noise. The typical sources of noise when operating in low-light conditions are considered in the generation of the datasets, namely, photon shot noise, circuit noise (including RTS noise) and dark current shot noise.

The extension of bilateral filter for sets of multiaperture images formulated in (4) preserves the classic smoothing parameters of a bilateral filter, in spatial and intensity domain: $\sigma_s, \sigma_i$, which are to be adjusted depending on the noise level. In order to provide an evaluation that is not negatively biased by a wrong parameter selection, we optimize them. The optimization is the minimization problem of finding the values of $\sigma_s, \sigma_i$ which lead to the minimal $l_2$ distance between the filtered image $\hat{\vec{I}}$ obtained from (4) using those parameters, and the corresponding noiseless original image (see Fig. 4). A set of optimal parameters is computed for each one of the three light levels considered, which correspond to three different levels of Poisson noise. In all cases the parameters are optimized by exhaustive search. This way, we can provide error manifolds showing the variability of the error with variations in the smoothing parameters. Two examples of these manifolds are given in Fig. 5, where the root mean square error (RMSE) between the filtered image and the original is plotted against $\sigma_s$ and $\sigma_i$, for both filtering domains. An absolute minimum for relatively low values of $\sigma_s$ and $\sigma_i$ is observed in both cases.

Regarding the alternative method for joint denoising of the multiaperture images, based on CS, there exist also several parameters to be adjusted. Probably the most relevant is the number of measurements to be performed per multiaperture image, which determines the number of rows of the sensing matrix, $m$, i.e., the number of linear equations to fulfill and, consequently, the quality of the reconstruction. Too few measurements might lead to information loss, while a

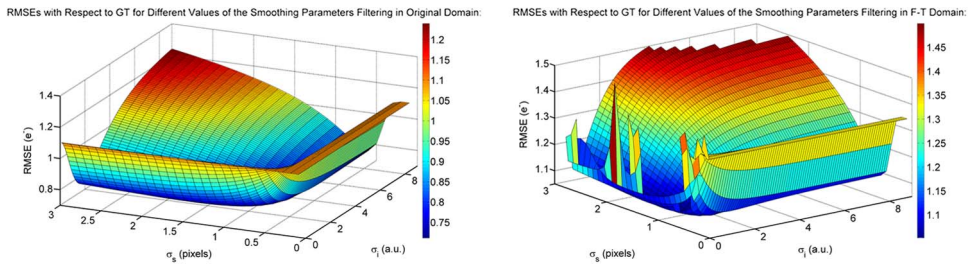



Fig. 5. Root mean square error (RMSE) between the image obtained with our joint filtering approach from the multiaperture set and the original image, which is used to generate the multiaperture dataset, for different values of the free parameters $\sigma_s, \sigma_i$. (Left) Results obtained when filtering in the original intensity domain. (Right) Results obtained when filtering in Freeman-Tukey (F-T) transformed domain. These results were obtained using the central 100 $\times$ 100 patch of the "lighthouse" images.

number of measurements too close to the dimensionality of the images might leave no freedom to the total variation minimization itself and lead to recovering the noise. In this paper we only present results obtained with $m = 0.8$ $n = 32000$ measurements per image, which means a 20% compression rate at sensing. Experimental results obtained using different values of $m$ are omitted for brevity. Nevertheless, higher number of measurements did not lead to any improvement in the denoised images. In order to efficiently solve the minimization in (6), we use the total variation minimization by augmented Lagrangian (TVAL) method of [37], which is implemented in the TVAL3 library. The solver accepts two penalty parameters, $\mu$ and $\beta$, that are to be adjusted according to the expected noise level in the images. The main parameter is the $\mu$ of (6), which is meant to establish a compromise between faithfulness to the measurements and TV-minimization. Like in the filtering approach, the parameters are optimized prior to evaluation. Nevertheless, the method is quite stable and the quality of the results does not degrade along quite large parameter ranges around the optimal point.

The results of our two methods for enhancing low light multiaperture images when applied to the synthetic datasets, namely "bridge," "chirp," and "lighthouse," are given in Figs. 6–8, respectively. All three figures are organized as follows: the three light level cases $(n_{e^-}^{\max} = \{1, 3, 9\})$ are grouped by rows, in ascendant order from top to bottom. The first column shows one of the raw multiaperture images, without any processing. The second column shows the result of the selective averaging method proposed in [13]. The third and fourth columns provide the results obtained applying our bilateral filtering extension in the original intensity domain and in the Freeman-Tukey (F-T) domain, respectively. The last column presents the result recovered by TV-minimization from compressed measurements.

#### 4.1.1.1. Bridge dataset

Observing Fig. 6, it becomes clear that the selective averaging method [column (b)] cannot cope with the extreme Poisson noise in the case of $n_{e^-}^{\max} = 1$ (first row). Both the filtering approaches and the CS-recovery face severe information loss due to noise in the raw data. Still, they produce results where the shadowed area under the bridge (black region), as well as the bridge fence (horizontal white strip), are distinguishable, still not clearly resolved. Unlike low-pass filtering, our methods are able to partially capture the structure of the small waterfall and the foliage of the background trees, completely lost in the selective averaging result. The result of filtering in F-T domain [see column (d)] seems to suffer from certain loss of the signal power, while the result of CS-recovery [see column (e)] is tessellated. In the case of $n_{e^-}^{\max} = 3$ (second row), all the results are clearly better and finer structure, e.g., the bridge fence, is recovered. While no more structure than in the selective averaging result is recovered, our methods are able to get rid of the noise without degrading the signal. Of special interest is the last row, corresponding to $n_{e^-}^{\max} = 9$, which is a more realistic case. The Poisson noise, which is strong in the raw images, cannot be completely eliminated by selective averaging but is completely filtered

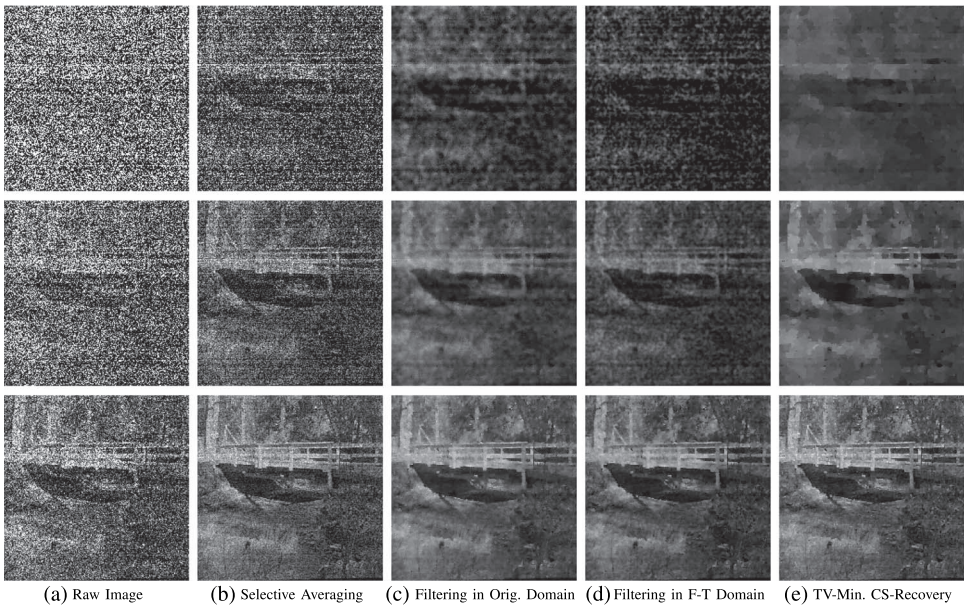



Fig. 6. Results of the low-light multiaperture image enhancement experiment for the "bridge" dataset. Three different levels of light are considered by rows from top to bottom: $n_{e^-}^{\max} = \{1, 3, 9\}$. The results are organized by columns as follows: raw multiaperture image (a), selective averaging result (b), filtering in original intensity domain (c), filtering in Freeman-Tukey (F-T) domain (d), and result of the compressive sensing (CS) recovery (e). These results are to be compared to the original scene on the left side of Fig. 4.

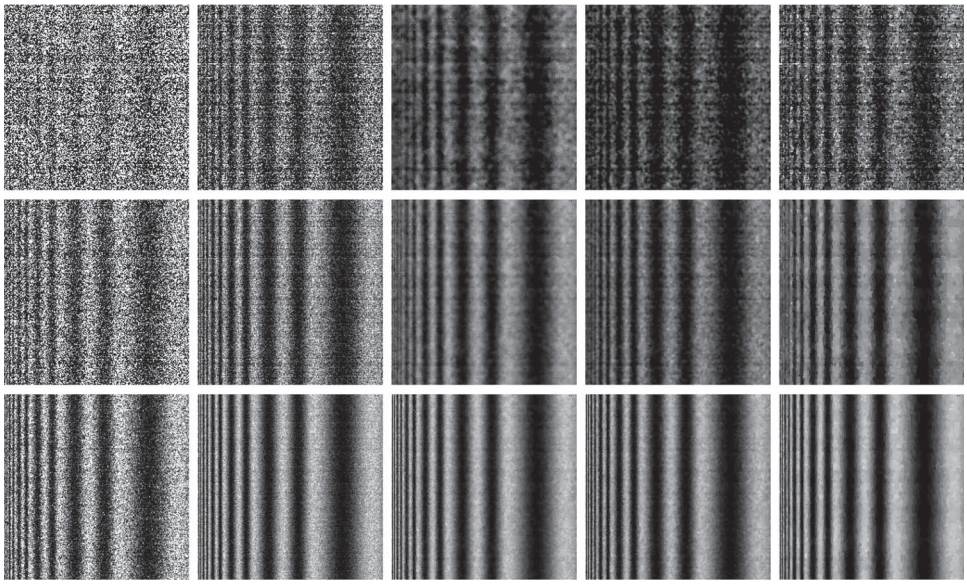
out by our methods, while keeping the finest details, e.g., foliage of the trees and small waterfall. Note the similarity between our three results. Provided that the filtering approaches and the CS approach are fundamentally different, the similarity of the results is an indicator of good performance of the methods.

#### 4.1.1.2. Chirp dataset

The dataset based on the middle image of Fig. 4 is intended to allow studying the frequency behavior of the different methods considered in this paper. The results in Fig. 7 follow the trend already observed in Fig. 6. Differently from Fig. 6, the results in the first row $(n_{e^-}^{\max} = 1)$ are surprisingly good and even relatively high frequencies are visible in the images. Confront, for instance, the original raw image (a) with the result of the filtering method in original domain (c). In the second row $(n_{e^-}^{\max} = 3)$, the selective averaging method seems to show less attenuation of the highest frequencies than our methods, still at the cost of higher noise levels. In the last row $(n_{e^-}^{\max} = 9)$, all methods seem to recover all frequencies equally well, being the proposed ones those showing better denoising capabilities.

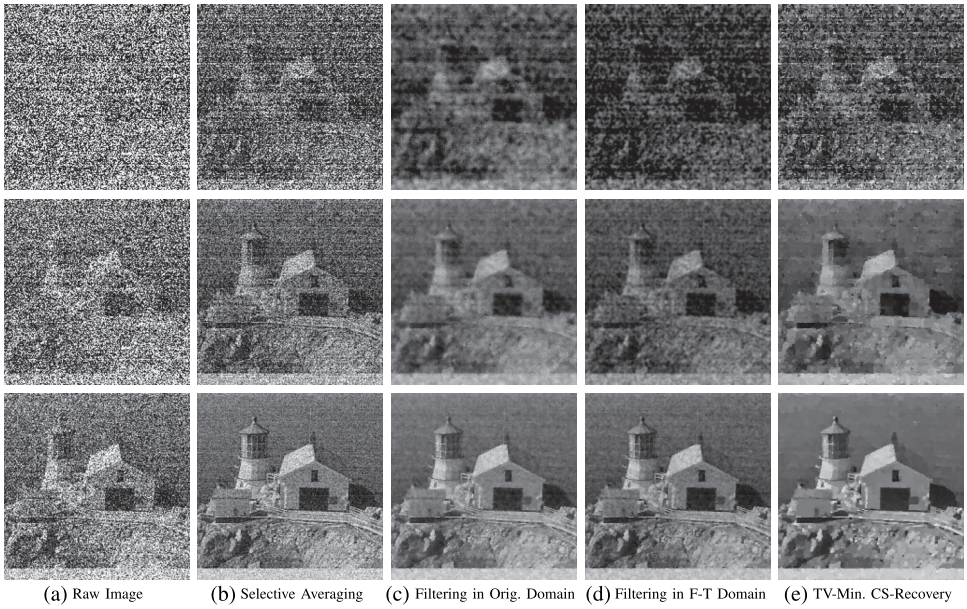
#### 4.1.1.3. Lighthouse dataset

The dataset based on the right side image of Fig. 4 is of special interest due to the diverse man-made structure in the scene, offering a large range of spatial frequencies, from very low (quasi-constant areas, e.g., the ocean or the walls) to extremely high (e.g., the antenna on top of the warehouse, already on the limit imposed by the Shannon criterion). Part of the main structure is recovered by our methods in the case of $n_{e^-}^{\max} = 1$ (see the first row of Fig. 8). Consider the fairly good result of filtering in original domain (c), where the lighthouse and the warehouse building are distinguishable, still not well-resolved. The rectangular main door and upper window can also be distinguished. In the case of $n_{e^-}^{\max} = 3$ (see the second row of Fig. 8), the antenna is still not distinguishable, but our methods are already able to approximately recover the



(a) Raw Image (b) Selective Averaging (c) Filtering in Orig. Domain (d) Filtering in F-T Domain (e) TV-Min. CS-Recovery

Fig. 7. Results of the low-light multiaperture image enhancement experiment for the "chirp" dataset. Three different levels of light are considered by rows from top to bottom: $n_{e^-}^{\max} = \{1, 3, 9\}$. The results are organized by columns as follows: raw multiaperture image (a), selective averaging result (b), filtering in original intensity domain (c), filtering in the Freeman-Tukey (F-T) domain, (d) and result of the compressive sensing (CS) recovery (e). These results are to be compared to the original scene in the middle part of Fig. 4.



(a) Raw Image (b) Selective Averaging (c) Filtering in Orig. Domain (d) Filtering in F-T Domain (e) TV-Min. CS-Recovery

Fig. 8. Results of the low-light multiaperture image enhancement experiment for the "lighthouse" dataset. Three different levels of light are considered by rows from top to bottom: $n_{e^-}^{\max} = \{1, 3, 9\}$. The results are organized by columns as follows: raw multiaperture image (a), selective averaging result (b), filtering in original intensity domain (c), filtering in the Freeman-Tukey (F-T) domain (d), and result of the compressive sensing (CS) recovery (e). These results are to be compared to the original scene on the right side of Fig. 4.

TABLE 1

Root mean square error (RMSE) of the low light multiaperture image enhancement results with respect to ground truth

| RMSE ($e^-$) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dataset Name | Bridge | | | Chirp | | | Lighthouse | | |
| Light Level | $n_{e^-}^{\max}=1$ | $n_{e^-}^{\max}=3$ | $n_{e^-}^{\max}=9$ | $n_{e^-}^{\max}=1$ | $n_{e^-}^{\max}=3$ | $n_{e^-}^{\max}=9$ | $n_{e^-}^{\max}=1$ | $n_{e^-}^{\max}=3$ | $n_{e^-}^{\max}=9$ |
| Raw | 10.798 | 10.842 | 10.952 | 10.798 | 10.845 | 11.002 | 10.799 | 10.845 | 11.004 |
| Simple Averaging | 6.6116 | 6.6210 | 6.6439 | 6.6132 | 6.6191 | 6.6494 | 6.6118 | 6.6233 | 6.6494 |
| Selective Averaging | 0.8567 | 0.9158 | 1.0799 | 0.8606 | 0.9348 | 1.1139 | 0.8615 | 0.9304 | 1.1195 |
| Filtering (Original) | 0.2205 | 0.3597 | 0.6930 | 0.2265 | 0.3744 | 0.5952 | 0.2184 | 0.3548 | 0.6525 |
| Filtering (F-T) | 0.3887 | 0.4341 | 0.9001 | 0.4194 | 0.6022 | 0.8073 | 0.3442 | 0.4588 | 0.8753 |
| CS-TV | 0.1556 | 0.3677 | 0.7581 | 0.3860 | 0.3834 | 0.5758 | 0.3514 | 0.3678 | 0.6876 |

TABLE 2

Peak signal-to-noise ratio (PSNR) of the low light multiaperture image enhancement results with respect to ground truth

| PSNR (dB) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dataset Name | Bridge | | | Chirp | | | Lighthouse | | |
| Light Level | $n_{e^-}^{\max}=1$ | $n_{e^-}^{\max}=3$ | $n_{e^-}^{\max}=9$ | $n_{e^-}^{\max}=1$ | $n_{e^-}^{\max}=3$ | $n_{e^-}^{\max}=9$ | $n_{e^-}^{\max}=1$ | $n_{e^-}^{\max}=3$ | $n_{e^-}^{\max}=9$ |
| Raw | -20.667 | -11.160 | -1.705 | -20.667 | -11.162 | -1.745 | -20.668 | -11.162 | -1.746 |
| Simple Averaging | -16.406 | -6.8761 | 2.6363 | -16.408 | -6.8735 | 2.6269 | -16.406 | -6.8790 | 2.6291 |
| Selective Averaging | 1.3435 | 10.307 | 18.417 | 1.3043 | 10.128 | 18.148 | 1.2956 | 10.169 | 18.104 |
| Filtering (Original) | 13.131 | 18.425 | 22.270 | 12.751 | 18.074 | 23.591 | 13.213 | 18.543 | 22.794 |
| Filtering (F-T) | 8.2060 | 16.792 | 19.999 | 7.5481 | 13.947 | 20.945 | 9.2633 | 16.311 | 20.242 |
| CS-TV | 16.171 | 18.233 | 21.490 | 8.2672 | 17.870 | 23.879 | 9.0825 | 18.229 | 22.338 |

silhouette of the chimney of the warehouse, masked by noise in the selective averaging result. The antenna is properly recovered by the filtering methods (c) and (d), as well as the selective averaging method (b) in the case of $n_{e^-}^{\max}=9$ (third row of Fig. 8). The poor PSNR of the selective averaging result may lead to confuse the antenna with background noise. The CS recovery result attenuates the antenna due to the implicit requirement of a sparse solution in gradient domain. Observe also the fence at the right of the warehouse, which is also one-pixel-wide and still recovered by all methods considered. Another critical area is that containing the central rocky cliff, where the noise in the selective averaging result can be erroneously interpreted as structure. The *noisy* points recovered by our methods in this area are not noise, but high-frequency structure contained in the original image (cf. the right side of Fig. 4).

The availability of ground truth allows computing errors for all the experimental results and performing a quantitative evaluation of the different methods for multiaperture image enhancement in low-light conditions. Two error values are computed for each of the images presented in Figs. 6–8: the root mean square error (RMSE) and the peak signal-to-noise ratio (PSNR). The RMSE represents the $l_2$ distance between the image and the ground truth, while the PSNR is a well-known parameter in photography, given by the ratio between the maximum signal power and the noise power, often in logarithmic scale. Tables 1 and 2 show the RMSE and PSNR, respectively, for all the results of the experiments with synthetic multiaperture datasets.

Tables 1 and 2 clearly show that the errors are very similar for all the three datasets, meaning that the performance of the different methods mostly depends on the low-light conditions, i.e., on the level of noise in the raw data, but not on the scene being acquired, confirming the reproducibility of the evaluation with different multiaperture datasets. The RMSE in the raw data is approximately around 11 in all cases considered, in all datasets. Simple averaging of the raw images reduces it to 6.6, while the selective averaging method is able to achieve RMSEs between 0.86 and 1.1, i.e., one order of magnitude lower than that of the raw images. The methods based on filtering and CS-recovery largely outperform the selective averaging method in terms of RMSE, especially in the case of lowest light $(n_{e^-}^{\max}=1)$, where the error reduction achieves a maximum of 98.56% with respect to the raw image and 81.84% with respect to the selective averaging result, in the “bridge” dataset.

Provided that the RMSE depends on the value of $n_{e^-}^{\max}$, a more independent indicator of the image quality is the PSNR, since it takes into account the maximum value that a pixel should deliver. A simple averaging brings an improvement of around 4.3 dB with respect to the raw images. The selective averaging method widely outperforms this, with an additional PSNR increments ranging between 15 and 18 dB, i.e., typically over 20 dB improvement with respect to the raw images. Our methods bring an appreciable improvement in all light cases considered, but especially large in the lowest light cases, where it is more needed. Our methods provide up to 37 dB PSNR improvement with respect to the raw images in the case of $n_{e^-}^{\max} = 1$, up to 30 dB in the case of $n_{e^-}^{\max} = 3$ and up to 26 dB in the case of $n_{e^-}^{\max} = 9$. Determining which method is the best is not an easy task, especially due to the very close performance of the method based on filtering in the original domain and the CS-based method. It seems that the filtering in the F-T-transformed domain performs slightly worse, especially in the lowest-light cases, while performing similarly to the other two methods for $n_{e^-}^{\max} = 9$. This observation is not a general rule and, e.g., in the case $n_{e^-}^{\max} = 1$ of the lighthouse dataset, filtering in the F-T domain outperforms the CS-based method.

#### 4.1.2. Evaluation of the Modulation Transfer Function

The "chirp" original image contains known spatial frequencies, increasing in horizontal direction from right to left. The exact spatial frequencies are $1/2^k$ cycles per pixel, where $k \in \mathbb{N}$, $k \in [1, 6]$. The different frequencies cover a region equal to its spatial period in horizontal dimension, also known beforehand. The Modulation Transfer Function (MTF) is a common mathematical descriptor of the contrast that an optical system or a camera is able to transmit or capture at a given spatial frequency. Consider the following general formula for the MTF:

$$\mathrm{MTF}(f) = \frac{\text{Degraded Image Contrast}(f)}{\text{Original Image Contrast}(f)} \tag{11}$$

where $f$ denotes spatial frequency. Optical systems and cameras exhibit MTFs that decay with $f$, i.e., low spatial frequencies are transmitted or captured without attenuation, while high frequencies are heavily attenuated. For each frequency in the "chirp" image, the denominator of (11) is obtained from the original image, while the numerator is obtained from the noisy image, after applying the different noise-reduction methods we consider in this paper. Since the width of each region of the "chirp" image is equal to the period of the corresponding spatial frequency, the second coefficient of a Fast Fourier Transform (FFT) of the vector obtained from averaging the image along the vertical direction directly provides the content of the desired frequency.

The MTF is calculated for the six frequencies contained in the "chirp" image in all the experimental cases considered, namely, simple averaging, selective averaging, filtering in original domain, filtering in F-T domain and CS recovery based on TV-minimization. The results are presented in three plots in Fig. 9, one for each light level considered $(n_{e^-}^{\max} = \{1, 3, 9\})$.

Simple averaging seems to offer a quite constant MTF, close to one, for all cases considered. For $n_{e^-}^{\max} = 1$, i.e., left plot of Fig. 9, the noise results in an MTF visibly greater than one for low frequencies $(f = 2^{-5}$ cycles/pixel$)$. The method of selective averaging preserves quite planar MTFs, while avoiding values greater than one thanks to its better noise-reduction capabilities. The MTFs for filtering in original domain and for the CS recovery are as good as those for selective averaging for low spatial frequencies, while decaying for high frequencies (as it is natural in optical systems). In other words, we establish a compromise between better denoising capabilities and a slightly worse frequency response in cases of very low-light. It is observable that the filtering in F-T domain has a strong low-pass filtering effect, which might degrade the signal for the lowest light level $(n_{e^-}^{\max} = 1)$. For $n_{e^-}^{\max} = 9$, the performance is close to those of the filtering in the original domain and the CS recovery.

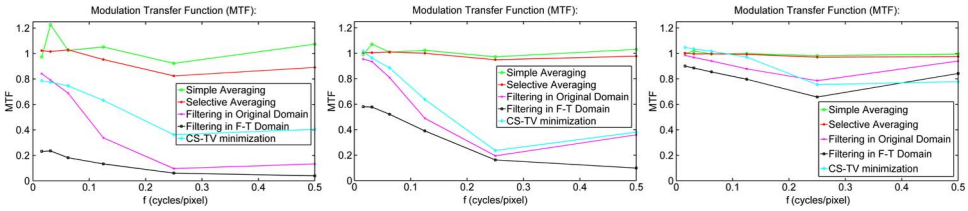


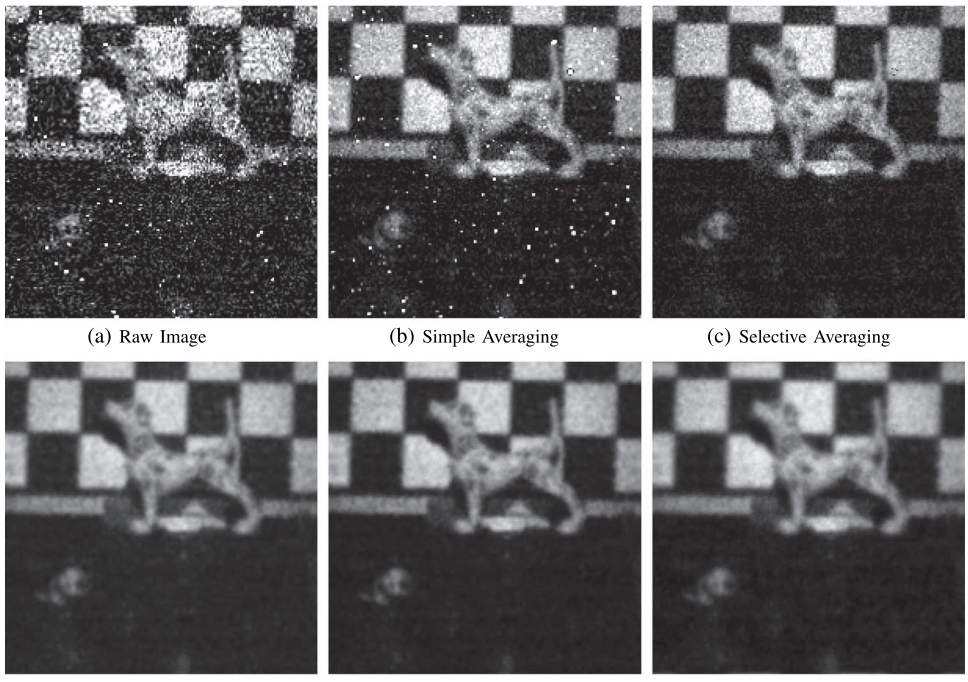

Fig. 9. Modulation transfer functions obtained from the "chirp" dataset. From left to right, the plots correspond to three different levels of light, given by the maximum number of electrons $n_{e^-}^{\max} = \{1, 3, 9\}$, respectively.

### 4.1.3. Qualitative Evaluation Using Real Data

In order to verify the good results obtained for synthetic data in a real system, we use the $3 \times 3$ multiaperture camera in [13] to gather real data. Each aperture has an $F$-number of 3 and provides images of $200 \times 200$ pixels. The sensor has a sensitivity of 20 V/lx s for light from a 3746 K light source [38]. The real acquisition is gathered under very low light. The scene is composed by a figure of a dog in front of a checkerboard. The illuminance on the checkerboard surface was measured to be $3 \times 10^{-2}$ lx. The number of electrons generated in the pixels sensing the white squares of the checkerboard (the brightest areas) is expected to be $n_{e^-}^{\max} = 9.3$ in absence of noise, i.e., similar to the highest light level considered in the synthetic datasets. The absence of ground truth does not allow computing errors and the evaluation of the results is made by visual inspection and, therefore, subjective. Since we cannot evaluate the quality of the results quantitatively, no parameter optimization is carried out. The parameters for the approach based on bilateral filtering are chosen to be $\sigma_s = 1$ and $\sigma_i = 1$. The intensity of the raw images is maximum-normalized and, therefore, takes values in [0,1]. These parameter values are conservative and higher values may be used if further denoising is required. The CS approach, based on TV-minimization, is carried out with the same sensing schema, i.e., $m = 0.8\, n = 32000$ measurements per image (20% compression), using binary sensing matrices with values $\phi_{i,j} \in \{-1, 1\}$, whose elements are drawn from a two-point distribution with equal probability for both cases. For the TV-minimization recovery, conservative parameter values are also chosen, namely $\mu = 8$ and $\beta = 8$. Further denoising can be achieved using different values. Decreasing $\mu$ reduces the noise at the eventual cost of blurring the image. Increasing $\beta$ reduces the noise at the cost of eventual tessellation of the image.

The results of the evaluation using real data are presented in Fig. 10. The first row shows one of the raw images and the results of state-of-the-art methods, such as simple averaging and selective averaging. The second row shows the results achieved with the low-light image enhancement methods for multiaperture systems presented in this paper.

The results in Fig. 10 confirm the superior performance of our methods when operating with real multiaperture images. As it was already observed in the quantitative error evaluations using synthetic datasets, presented in Tables 1 and 2, the performance of the filtering method and CS-recovery are similar. The use of a bilateral filter framework to combine the raw images provides enhanced robustness to spurious intensity values. Consider, for instance, the bottom-left corner of the white checkerboard square next to the dog's tail. There are few pixels with a visibly low value (black pixels) in Fig. 10(c). Confront it with the filtering results [see Fig. 10(d) and (e)], where these pixels exhibit acceptable (close to white) values. Even the CS-recovery [Fig. 10(e)], which uses the selective averaging method to generate a valid starting point for the TV-minimization, achieves a better estimation for these pixels than the selective averaging method. Similar cases can be observed in other areas of the image, e.g., the dog's body, for which our methods provide a cleaner result. Another region of interest is the dark area in the lower half of the image. Filtering in the original domain does not bring much further denoising with respect to the selective averaging method, while filtering in the F-T domain seems to deliver a better



(a) Raw Image (b) Simple Averaging (c) Selective Averaging

(d) Filtering in Original Domain (e) Filtering in F-T Domain (f) TV-Minimization CS-Recovery

Fig. 10. Results of the low-light multiaperture image enhancement experiment using real data. The light intensity on the background checkerboard was $3 \times 10^{-2}$ lx. The first row shows one of the raw images (a) and the results of simple averaging (b) and selective averaging (c) of the raw images. In the second row, we present the results of combining the raw data using our methods for low-light image enhancement, namely, the filtering approach in original domain (d), Freeman-Tukey (F-T) domain (e), and the CS-recovery from stochastic measurements (f).

solution. The best method for this area seems to be the CS-recovery, probably because the black region perfectly meets the hypothesis that the optimal solution is the one of minimal TV.

### 4.2. Multitap Camera

We evaluate the method described in Section 3.3 with raw data from a real multitap camera. We use the multimodal sensor ZESS MultiCam [30], equipped with the medium range IR illumination system presented in [39], which provides an uniform illumination along a fairly large field of view (FOV). The MultiCam features exchangeable optics, and therefore, the FOV of the camera can be adjusted by choosing an appropriate lens. In our case, a 8.5 mm lens is used, leading to a FOV of $46^\circ \times 35^\circ$. The MultiCam is a multimodal sensor featuring both a color camera and a depth sensor. The use of a single lens for both is possible by means of a Bauernfeind prism with an integrated beam splitter. Since the color modality is not considered in our methods, we focus on the depth imaging hardware. The depth sensing array is a PMD 19K-S3 [40], with an image size of $120 \times 160$ pixels. Each pixel is a 2-tap pixel, i.e., there are two integration channels, often referred to as A and B. Provided that the PMD 19K-S3 operates according to the four phases algorithm [see (1)], four sequential acquisitions are required, at four equally-spaced phase shifts of the binary signal that controls the integration. This means, a set of $N = 8$ multitap images are needed to generate the depth image. Such sets of images are the input data of our method.

Regarding the experimental setup, a scene is created with depths ranging from 1.5 to 2.0 m, approximately. It contains objects of known geometry, such as two balls of different sizes on a table. The camera looks frontally to a plain wall. At the right side a panel is placed, parallel to the wall and closer to the camera. The panel contains two Böhler stars [41], which are the 3-D version of the widely-used Siemens stars. The stars are used to detect an eventual degradation

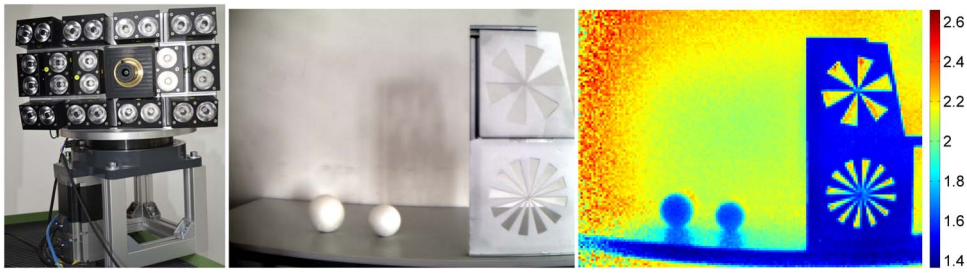



Fig. 11. The first image from the left is our medium-range ToF imaging system. The camera, surrounded by LED modules, is a ZESS MultiCam, which delivers both RGB and ToF depth images. The use of a monocular setup avoids parallax-related registration errors. The modules are driven synchronously and oriented in order to get an uniform and complete IR illumination of the scene. The central and right images are the color and depth images of our experimental setup, as observed by the MultiCam. The depth scale is in meters.

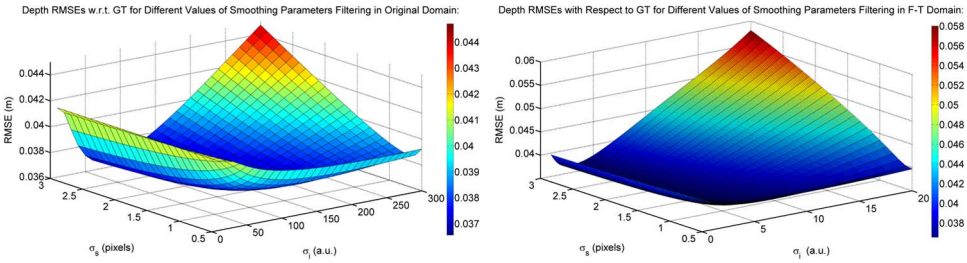



Fig. 12. Root mean square error (RMSE) between the depth image obtained using filtered raw data and a reference depth image, taken as ground truth (GT), for different values of the free parameters, $\sigma_s, \sigma_i$. The left plot shows the results obtained when filtering in the original intensity domain and the right plot when filtering in Freeman-Tukey transformed domain (F-T). The original raw data was acquired with 50 $\mu$s exposure time.

of the angular resolution in the final depth image. Fig. 11 provides the typical output of the MultiCam when recording the scene in presence of some ambient light and using sufficiently large exposure times.

The datasets are acquired using abnormally low exposure times, which are 2 to 3 orders of magnitude lower than those required for an appropriate sensing of the scene, in the millisecond range. Three different exposures are considered: 10 $\mu$s, 20 $\mu$s, and 50 $\mu$s. For 10 $\mu$s, the scene is hardly visible in the raw images and the signal level close to the noise level of the sensor.

There exist two free parameters to adjust in the filter proposed in Section 3.3, namely, $\sigma_\mathrm{s}$ and the single value assigned to all diagonal elements of $\Pi_{\vec{s}}$, $\sigma_\mathrm{i}$. In order to assure that our results are not affected by a wrong parameter selection, they are optimized within a certain feasible range. The optimization is carried out by exhaustive search and the cost function to minimize is the distance between the depth image obtained from the filtered raw data and a reference depth image $\vec{d}_\mathrm{GT}$, obtained using a long exposure time (see the right side of Fig. 11). For clarity, the optimization problem is formulated as

$$[\sigma_\mathrm{s}, \sigma_\mathrm{i}] = \underset{\substack{\sigma_{\mathrm{s}_\mathrm{min}} \le \sigma_\mathrm{s} \le \sigma_{\mathrm{s}_\mathrm{max}} \\ \sigma_{\mathrm{i}_\mathrm{min}} \le \sigma_\mathrm{i} \le \sigma_{\mathrm{i}_\mathrm{max}}}}{\operatorname{argmin}} \left\| \vec{d}\left(\hat{\boldsymbol{S}}(\sigma_\mathrm{s}, \sigma_\mathrm{i})\right) - \vec{d}_\mathrm{GT} \right\|_2 \tag{12}$$

where $\hat{\boldsymbol{S}}(\sigma_\mathrm{s}, \sigma_\mathrm{i})$ is the matrix composed by the filtered multitap images, stacked by columns, where $\sigma_\mathrm{s}, \sigma_\mathrm{i}$ is the parameter values used in the filtering. $\vec{d}(\hat{\boldsymbol{S}}(\sigma_\mathrm{s}, \sigma_\mathrm{i}))$ is the final depth image, computed from the filtered images using Eq. (1). Obviously, the optimal parameters are different in the original intensity domain and in the F-T transformed domain. Fig. 12 provides the error manifolds obtained from the optimization for the 50 $\mu$s exposure case, in both domains. The

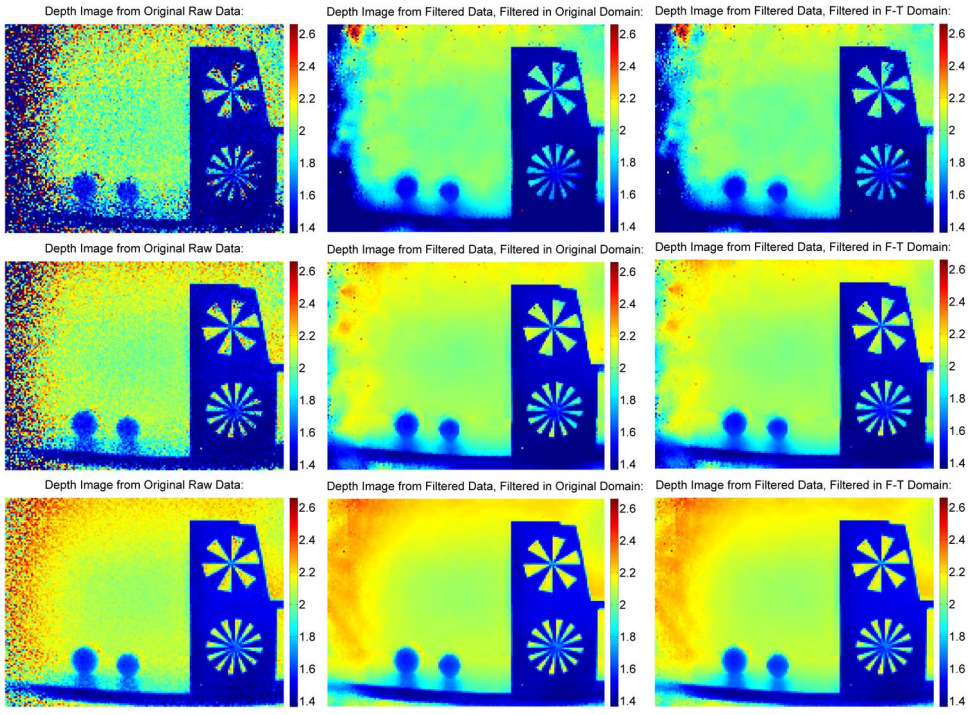



Fig. 13. Results of the low-light multitap image enhancement experiment. The images are the final depth images obtained using very short exposure times, namely, 10 $\mu$s (first row), 20 $\mu$s (second row), and 50 $\mu$s (third row). For each exposure case, three results are given, from left to right: depth image computed from the original raw data, depth image computed from raw data filtered in the original intensity domain, and depth image computed from raw data filtered in the Freeman-Tukey (F-T) transformed domain. Scales are in meters.

considered parameter domain, $[\sigma_{s_{min}}, \sigma_{s_{max}}] \times [\sigma_{i_{min}}, \sigma_{i_{max}}]$, was chosen to be [0.5, 3.0] $\times$ [10, 300] in the original domain and [0.5, 3.0] $\times$ [1.0, 20.0] in the F-T domain.

The results of the experiments, obtained using the optimal filter parameters for the different exposure cases, are presented in Fig. 13 in the shape of depth images. The first column shows the depth images obtained from the original raw data, while the second and third columns show the depth images obtained from jointly filtered data, in original and F-T domains, respectively. These results are to be compared the right side of Fig. 11, which was taken as reference depth map of the scene.

For all experimental cases considered, the depth images obtained from filtered data exhibit visibly better quality than those obtained directly from the raw data. Far from being over-smoothed by the filter, depth gradients are preserved or even enhanced (compare the fields of the upper star in the depth image obtained from original raw data to those of the depth images obtained from filtered data). In the 10 $\mu$s case (first row), the round shape of the balls becomes visible. In the 20 $\mu$s case (second row), the table surface is clearly recovered and depth estimation becomes possible also for the left part of the image, under poorer illumination conditions. Finally, in the 50 $\mu$s case (third row), the visual quality of the depth images obtained from filtered data is so high that they might be considered even better than the reference image on the right side of Fig. 11. In general, the Böhler stars indicated that the filtering does not lead to loss of angular resolution in the depth images. In order to provide a quantitative evaluation of the improvement achieved by filtering prior to depth calculus, we compute the root mean square error (RMSE) between the results in Fig. 13 and the reference depth map on the right side of Fig. 11. The results of this error evaluation are given in Table 3.

The depth RMSEs are coherent with the visual quality of the depth images in Fig. 13. The proposed method leads to an approximate depth error reduction of 50% for 10 $\mu$s exposure time,

TABLE 3

Root mean square error (RMSE) of the depth images in Fig. 13 with respect to ground truth (GT; see the right side of Fig. 11)

| | RMSE (cm) | | |
|---|---|---|---|
| Exposure Time ($\mu s$) | | Filtering Domain | |
| | No Filter | Original Intensity | Freeman-Tukey |
| 10 | 64.46 | 31.86 | 32.14 |
| 20 | 36.99 | 12.30 | 12.37 |
| 50 | 9.02 | 5.37 | 5.39 |

77% for 20 $\mu$s, and 40% for 50 $\mu$s. We observe that the depth error is independent from the domain where the filtering was carried out, being the filtering in the original intensity domain the one delivering slightly better results.

## 5. Conclusion

In this paper, we deal with low-light imaging using multiaperture and multitap systems. A common characteristic of these systems is that several images of the same scene are acquired, either simultaneously or within a short time. In the case of conventional multiaperture systems, all the images are acquired simultaneously and expected to be identical, while in the case of multitap systems, where the integration is governed by a control signal, the images are acquired sequentially and expected to differ from each other. Multiaperture multitap systems with coded apertures allow simultaneous acquisition. We show that, for all these systems, a bilateral filter framework can be adapted to exploit the information redundancy, while efficiently filtering out the noise. We consider filtering in the original pixel domain and in a transformed domain, obtained through a variance-stabilizing transform, in order to normalize the Poisson noise. We also show that a compressive sensing (CS) framework can be used as well for exploiting the information redundancy in the case of pure multiaperture systems. Our experiments show that a denoised image can be recovered from few measurements of the raw images, highly corrupted by Poisson noise, imposing sparsity in the gradient domain.

The performance of our methods for the multiaperture case is evaluated both with synthetic and real data from an experimental multiaperture camera. Experiments using different synthetic datasets confirm that our methods can achieve significantly higher error reduction than state-of-the-art methods based on selective averaging of the raw images. This improvement is especially large in the cases of lowest light, where selective averaging cannot cope with the overwhelming levels of noise. For instance, for one of the datasets we register a 37 dB PSNR improvement with respect to the raw images in the case of a maximum number of $n_{e^-}^{\max} = 1$ electron is expected per pixel, which is 15 dB higher than that achieved by selective averaging. The CS approach performs as well as the best filtering approach. Filtering in the variance-stabilized domain delivers worse results than filtering in the original domain in the cases of lowest light conditions, probably due to the poor stabilization capabilities of the transformation when the expected value of the original Poisson distribution approaches to zero.

A study of the modulation transfer function (MTF) reveals that our methods might attenuate the high frequencies, while offering better behavior than the selective averaging for medium and low frequencies. Since the high-frequency details are often masked by the intense noise in low-light imaging, the information loss due to high-frequency attenuation is largely compensated by the best low-frequency transfer, resulting in lower overall error with respect to ground truth. Filtering in the transformed domain typically leads to poor MTFs, especially in the lowest light cases.

As an application teaser, the performance of the filtering framework adapted to the multitap case is evaluated using data from a real PMD ToF camera. PMD ToF cameras estimate the depth from several sequential acquisitions of a 2-tap image sensor. Our experiments show that the proposed framework allows ToF depth imaging with low illumination or short exposure times.

The redundancy among raw images is used to remove uncorrelated noise without degrading the effective angular resolution. This results in a large noise reduction in the final depth images. Depth error reductions up to 77% have been observed. Exposure times can be decreased up to 2 or 3 orders of magnitude, enabling operation in the microsecond range.

## Appendix
## Proof of Equation (10)

Recall that (10) is to be derived from (9), provided that

$$\hat{x}(y) = \lim_{\left(t=\frac{y}{\sqrt{n+1}}\right)_{n\to\infty}} n\,\hat{p}(t).$$

We substitute $\hat{p}(t)$ in the previous expression by (9) and make use of the trigonometric limit approximation $\lim_{\varepsilon\to 0}\sin\varepsilon \simeq \varepsilon - (\varepsilon^3/6)$, which derives from the corresponding Taylor expansion, neglecting the terms of order equal or higher than five. We obtain the following polynomial:

$$\hat{x}(y) = \lim_{\left(t=\frac{y}{\sqrt{n+1}}\right)_{n\to\infty}} \frac{n}{2}\left[1-\sqrt{1-\left(\frac{n\,t+\frac{5t}{6}-\frac{1}{t}}{n}\right)^2}\right].$$

Note that the signature of the cosine has been left away, since in the Poisson case: $\lim_{\varepsilon\to 0}\cos\varepsilon = 1$. At this point, we get rid of the square root by using the limit approximation $\lim_{\varepsilon\to 0}\sqrt{1-\varepsilon} \simeq \varepsilon - (\varepsilon/2)$, which derives, in turn, from the first order Taylor expansion. Substituting and operating, we get

$$\hat{x}(y) = \lim_{\left(t=\frac{y}{\sqrt{n+1}}\right)_{n\to\infty}} \frac{1}{4}\left(\sqrt{n}\,t+\frac{5t}{6\sqrt{n}}-\frac{1}{t\sqrt{n}}\right)^2.$$

Executing the implicit changes of variables $t = y/\sqrt{n+1}$, we get a function of the transformed variable $y$

$$\hat{x}(y) = \lim_{n\to\infty} \frac{1}{4}\left(\frac{\sqrt{n}}{\sqrt{n+1}}\,y+\frac{5}{6\sqrt{n}\sqrt{n+1}}\,y-\frac{\sqrt{n+1}}{\sqrt{n}}\,y^{-1}\right)^2.$$

The limit can be now trivially calculated, leading to

$$\hat{x}(y) = \left(\frac{y-y^{-1}}{2}\right)^2.$$

■

## Acknowledgment

The authors would like to thank Prof. S. Kawahito for providing valuable real data gathered with the CMOS image sensor that he developed. M. H. C. would also like to thank Dr. A. Seel (ZESS) for his help to achieve a simple inversion formula for the Freeman-Tukey transformation in the Poisson case.

## References

[1] B. Wilburn *et al.,* "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005.
[2] K. Venkataraman *et al.,* "PiCam: An ultra-thin high performance monolithic camera array," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 166:1–166:13, Nov. 2013.

[3] F. Mochizuki *et al.,* "6.4 single-shot 200 mfps 53-aperture compressive CMOS imager," in *Proc. IEEE ISSCC*, 2015, pp. 1–3.
[4] J. Tanida *et al.,* "Thin observation module by bound optics (TOMBO): Concept and experimental verification," *Appl. Opt.*, vol. 40, no. 11, pp. 1806–1813, Apr. 2001.
[5] T. Spirig and P. Seitz, "*Vorrichtung und verfahren zur detektion und demodulation eines intensitätsmodulierten strahlungsfeldes*," Patent wO Patent App. PCT/EP1995/004,235, May 23, 1996. [Online]. Available: https://www.google.com/patents/WO1996015626A1?cl=un
[6] T. Spirig, *Smart CCD/CMOS Based Image Sensors With Programmable, Real-Time, Temporal and Spatial Convolution Capabilities for Applications in Machine Vision and Optical Metrology*, 1997. [Online]. Available: https://books.google.de/books?id=hK3QNwAACAAJ
[7] R. Lange, "3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. Siegen, Siegen, Germany, 2000. [Online]. Available: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2006/178/pdf/lange.pdf
[8] T. Möller *et al.,* "Robust 3D measurement with PMD sensors," in *Proc. 1st Range Imaging Res. Day ETH*, 2005, pp. 906463–906467.
[9] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sens. J.*, vol. 11, no. 9, pp. 1917–1926, Sep. 2011.
[10] B. Langmann, K. Hartmann, and O. Loffeld, "Depth camera technology comparison and performance evaluation," in *Proc. ICPRAM*, 2012, pp. 438–444.
[11] A. Payne *et al.,* "7.6 A 512 × 424 CMOS 3D time-of-flight image sensor with multi-frequency photo-demodulation up to 130 MHz and 2 GS/s ADC," in *Proc. IEEE ISSCC*, Feb. 2014, pp. 134–135.
[12] C. Leyris *et al.,* "Impact of random telegraph signal in CMOS image sensors for low-light levels," in *Proc. 32nd ESSCIRC*, Sep. 2006, pp. 376–379.
[13] B. Zhang *et al.,* "RTS noise and dark current white defects reduction using selective averaging based on a multi-aperture system," *Sensors*, vol. 14, no. 1, pp. 1528–1534, Jan. 2014.
[14] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.
[15] G. Petschnigg *et al.,* "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664–672, Aug. 2004.
[16] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007. [Online]. Available: http://doi.acm.org/10.1145/1276377.1276497
[17] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[18] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
[19] P. Besbeas, I. De Feis, and T. Sapatinas, "A comparative simulation study of wavelet shrinkage estimators for Poisson counts," *Int. Statist. Rev./Revue Internationale de Statistique*, vol. 72, no. 2, pp. 209–237, 2004.
[20] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[21] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
[22] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[23] J. Jin, B. Yang, K. Liang, and X. Wang, "General image denoising framework based on compressive sensing theory," *Comput. Graph.*, vol. 38, pp. 382–391, 2014.
[24] S. Muthukrishnan, "Data streams: Algorithms and applications," in *Proc. 14th Annu. ACM-SIAM Symp. Discrete Algorithms*, Philadelphia, PA, USA, 2003, pp. 413–413.
[25] Y. Zhang, "On theory of compressive sensing via l1-minimization: Simple derivations and extensions," CAAM, Rice Univ., Houston, TX, USA, TR08-11, Jul. 2008.
[26] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[27] M. Heredia Conde, K. Hartmann, and O. Loffeld, "A compressed sensing framework for accurate and robust waveform reconstruction and phase retrieval using the photonic mixer device," *IEEE Photon. J.*, vol. 7, no. 3, pp. 1–16, Jun. 2015.
[28] B. Langmann, K. Hartmann, and O. Loffeld, "Comparison of depth super-resolution methods for 2D/3D images," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 3, pp. 635–645, 2011.
[29] T. D. A. Prasad, K. Hartmann, W. Weihs, S. E. Ghobadi, and A. Sluiter, "First steps in enhancing 3D vision technique using 2D/3D sensors," in *Proc. CVWW*, O. Chum and V. Franc, Eds. Prague, Czech Republic, Feb. 2006, pp. 82–86.
[30] K. Hartmann and R. Schwarte, "*Detection of the phase and amplitude of electromagnetic waves*," U.S. Patent 7 238 927 B1, Jul. 3, 2007. [Online]. Available: https://www.google.com/patents/US7238927
[31] M. Heredia Conde, K. Hartmann, and O. Loffeld, "Structure and rank awareness for error and data flow reduction in phase-shift-based ToF imaging systems using compressive sensing," in *Proc. IEEE 3rd Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar Remote Sens.*, Jun. 2015, pp. 144–148.
[32] M. S. Bartlett, "The square root transformation in analysis of variance," *Supplement J. Roy. Statist. Soc.*, vol. 3, no. 1, pp. 68–78, 1936.
[33] F. J. Anscombe, "The transformation of Poisson, binomial, and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, 1948.
[34] M. F. Freeman and J. W. Tukey, "Transformations related to the angular and the square root," *Ann. Math. Statist.*, vol. 21, no. 4, pp. 607–611, 1950.
[35] A. Foi, *Optimization of Variance-Stabilizing Transformations*, 2009. [Online]. Available: http://www.cs.tut.fi/foi
[36] J. J. Miller, "The inverse of the Freeman Tukey double arcsine transformation," *Amer. Statist.*, vol. 32, no. 4, pp. 138–138, 1978.

[37] C. Li, "*Compressive sensing for 3D data processing tasks: Applications, models and algorithms*," Ph.D. dissertation, Dept. Comput. Appl. Math., Rice Univ., Houston, TX, USA, 2011, aAI3524544.
[38] M.-W. Seo *et al.,* "A low-noise high-dynamic-range 17-b 1.3-megapixel 30-fps CMOS image sensor with column-parallel two-stage folding-integration/cyclic ADC," *IEEE Trans. Electron Devices*, vol. 59, no. 12, pp. 3396–3400, Dec. 2012.
[39] B. Langmann, K. Hartmann, and O. Loffeld, "Real-time image stabilization for ToF cameras on mobile platforms," in *A State-of-the-Art Survey on Time-of-Flight and Depth Imaging: Sensors, Algorithms, and Applications*, vol. 8200, Lecture Notes in Computer Science, R. K. A. K. M. Grzegorzek and C. Theobalt, Eds. New York, NY, USA: Springer-Verlag, 2013, pp. 289–301.
[40] "*PMD PhotonICs 19k-S3*," 2014, [Online accessed Mar. 21, 2014]. [Online]. Available: http://pmdtec.com/html/pdf/pmdPhotonICs_19k_S3.pdf
[41] W. Böhler, V. M. Bordas, and A. Marbs, "Investigating laser scanner accuracy," in *Proc. 14th CIPA Symp.*, O. Atlan, Ed., Antalya, Turkey, Oct. 2003, vol. 34, pp. 696–701. [Online]. Available: http://cipa.icomos.org/fileadmin/template/doc/antalya/189.pdf