

# Compressive sensing computational ghost imaging

Vladimir Katkovnik and Jaakko Astola

Department of Signal Processing, Tampere University of Technology,

P. O. Box 553, Tampere, Finland, [vladimir.katkovnik@tut.fi](mailto:vladimir.katkovnik@tut.fi).

The computational ghost imaging with a phase spatial light modulator (SLM) for wave field coding is considered. A transmission mask amplitude object is reconstructed from multiple intensity observations. Compressive techniques are used in order to gain a successful image reconstruction with a number of these observations (measurement experiments) which is smaller than the image size. Maximum likelihood style algorithms are developed, respectively for Poissonian and approximate Gaussian modeling of random observations. A sparse and overcomplete modeling of the object enables the advanced high accuracy and sharp imaging. Numerical experiments demonstrate that an approximative Gaussian distribution with an invariant variance results in the algorithm which is quite simple in implementations and nevertheless efficient for Poissonian observations. © 2012

Optical Society of America

*OCIS codes:* 070.2025, 100.2000, 100.3010, 100.3190

## 1. Introduction

In the chaotic pseudothermal light ghost imaging (GI) a laser beam illuminates an transmission mask object and a transmitted light is collected by a single-pixel (bucket) sensor with no spatial resolution. The reconstruction of the object transmission is formed by correlating the bucket sensor's output with the output from a spatially resolving sensor that is illuminated by a reference beam correlated with the signal beam but has not interacted with the object.

There is a long list of publications that run the gamut from various experimental realizations to discussions of fundamental classical and quantum physics behind GI. Initially, it is demonstrated by Pittman et al. [1] that GI is enabled by specifically quantum effects. The discovery, that GI could be performed with the pseudothermal light, proved that GI admits the classical (coherence theory and statistical optics) and quantum-mechanical (nonlocal two-photon interference) interpretations. The both interpretations have been successfully used to analyze the thermal-light GI and to predict experimental observations [2, 3]. It has been demonstrated that the pseudothermal GI results from correlation of the object and reference fields [4, 5]. A comprehen-

sive review of the fundamentals of GI using classical and quantum physics is given by Erkmen and Shapiro [6] and Shih [7].

Based on the speckle-correlation interpretation of the thermal-light GI, Shapiro proposed to implement GI using a phase spatial light modulator (SLM) illuminated by a spatially coherent laser beam [8]. This development is a starting point of *computational GI*, where a reference wave field and high-resolution sensor are not compulsory because the reference speckle wave field and the sensor output can be precompute using the models for the free space wave field propagation. The first experimental proof of this idea was done by Katz et al. [9].

Further works in this field concern application and development of novel computational techniques. Originally, the computational GI is founded on calculation of the cross-correlation between the data collected by the high-resolution sensor in the reference arm and by the single-pixel sensor in the object arm. Contrary to it, *compressive sensing ghost imaging (CS-GI)* is based on *modeling and sparse approximations* of the transmission mask object. Then, the algorithms and reconstructions are solutions of special constrained optimization problems. The cross-correlations appear as implicit

parts in these algorithms.

The first work on CS-GI is done by Katz et al. [10], where the  $l_1$ -norm of the image spectral sparse approximation is used as an objective function of optimization. Another CS-GI technique is developed by Jiying et al. [11], where the maximum likelihood style approach is developed for noisy data.

*Compressive sensing* (CS) is a recent trend in image/signal processing having the potential to significantly extend the capabilities of digital imagers. The CS theory states that if the image/signal is *sparse*, then under certain conditions on the observation operator, it can be recovered exactly from the compressive measurements [12, 13]. The sparsity in CS assumes that there exists a basis consisting of a comparatively small number of items where the image can be represented exactly or approximately with high accuracy. This *ideal sparse* basis is a priori unknown and selected from a given set of potential bases (dictionary or dictionaries). The popularity and success of *CS* are due to the attractive theory, the efficient algorithms and the evidence that the developed formalism fits perfectly to many important applications. In CS the image is reconstructed from subsampled data. A total number of available observations can be smaller (much smaller) than size of the image. CS can be

treated as a special regularization technique for variational inverse imaging where regularization is introduced through a *sparse* and *overcomplete* object modeling.

Recently in optics, *sparse* and *overcomplete imaging* has become a subject of multiple applications, in particular, in the context of CS. Complex-valued signals and operators are distinctive features of this development. Basic facts of the corresponding theory, algorithms, simulations as well as experimental demonstrations are presented by Gazit et. al [14], where the sparsity constraints are exploited for sub-wavelength imaging overcoming the diffraction limitations. An extensive survey of the CS theory and techniques addressed to applications in optics is done by Willett et al. [15]. It is shown in the papers [10,11] that CS algorithms allow to get good reconstructions for GI provided that the total number of observations is much smaller than the number of pixels of the image. Contrary to it, in order to achieve a good quality by the cross-correlation methods the total number of observation should be much larger than the image size (e.g. [16]).

In this paper based on the recent works of our research group [17,18] we propose and develop the original CS algorithms which are able to give opti-

mal or near optimal object reconstructions due to a constrained maximum likelihood formulation of the problem. It is assumed that the observations are Poissonian, i.e. photon counting effects are taken in consideration. The proposed algorithms are iterative with decoupling of the inverse of the image formation model in observations and the filtering of errors in the object transmission. The sparse and overcomplete modeling of the object to be reconstructed is a special points of these algorithms. A family of the Block Matching 3-D (BM3D) algorithms for various imaging problems has been proposed within the framework of nonlocal patch-wise image modeling [19]. The analysis and synthesis operations from BM3D algorithms have been used in order to design the BM3D frames [17]. In this paper we use this advanced overcomplete BM3D image modeling for the object to be reconstructed.

While our approach is based on multi-objective optimization and essentially different from the conventional variational techniques using a single-objective optimization it is useful to provide references on some recent works on inverse imaging for Poissonian observations in CS. Some of the basic principles discussed in these works can be tracked in our technique. In the paper by Figueiredo and Bioucas-Dias [20] the imaging is formulated as the

constrained maximum likelihood optimization solved using the alternating direction method of multipliers. The technique uses splitting variables essentially beneficial for solution of partial optimization problems. The optimization formulation considered by Harmany et al. [21] exploits a penalized likelihood objective function and especially focussed on nonnegativity constraints for optical problems. It is demonstrated that the accurate treatment of these constraints may result in significant improvement of imaging in comparison with the crude projections of estimates on the quadrants of negative values. CS performance bounds under Poisson observations are studied by Raginsky et al. [22].

The rest of the paper is organized as follows. In Section 2 a mathematical modeling for observation formation in GI system is presented. It includes the Poissonian and approximative Gaussian observation models as well as the sparse and overcomplete modeling of the object transmission mask. The variational approach to the object reconstruction in GI is given in Section 3, where the algorithm derived for the Gaussian approximation with invariant standard deviation is presented. The derivation of the algorithms for Poissonian data and Gaussian approximation with varying standard deviations

are given in Appendices A and B. Simulation experiments are discussed in Section 4.

## 2. Observations and sparse overcomplete object modeling

The two-arm thermal-light GI setup with SLM is shown in Fig.1. The speckle wave field testing the object is generated by the monochromatic laser beam illuminating SLM used for modulation of the wave field. The reference (horizontal) and object (vertical) light beams are outputs from a 50-50 beam splitter. These beams undergo paraxial diffraction over free-space paths of length  $d$ , yielding identical (at least in theory) speckle measurement-plane fields  $u_r$ . The reference beam is registered by a high-resolution digital sensor with the beam intensity output  $|u_r|^2$ . The signal beam illuminates the transmission-mask object  $u_0$ , located immediately in front of the bucket (single-pixel) sensor. The bucket sensor measures the total intensity  $o_r$  of the wave field illuminating this sensor.

The single-arm computational setup of GI is considered in the form shown in Fig.2, where the precalculated intensity of the speckle field  $|u_r|^2$  is used instead of  $|u_r|^2$  measured in the reference arm of the two-arm GI setup. In the both setups, single- and two-arms, the object amplitude transmission

is reconstructed using the multiple outputs obtained for experiments with various phase distributions of SLM.

In this section we derive the image formation model linking observation intensities at the output of the single-pixel bucket sensor with the squared magnitudes of the transmission mask object and intensities of the speckle wave field generated by SLM. It is assumed that the wave field is coherent, the wave front is flat and the paraxial approximation can be used for wave field propagation. The derived models are applicable for the both single- and two-arm setups.

### *2.A. Observation modeling*

Let  $u_1(x)$  and  $u_2(x)$ ,  $x \in \mathbb{R}^2$ , denote complex-valued wave fields in the initial and following parallel planes, with a distance  $d$  between the planes. In discrete modeling the continuous argument  $x$  is replaced by the digital one with a corresponding replacement of all continuous functions by their discrete sampled counterparts:  $u_1(x) \rightarrow u_1[k]$ ,  $u_2(x) \rightarrow u_2[k]$  with  $2D$  integer argument  $k$ .

In what follows we use a vector-matrix notation with  $N_1 \times N_2$  images given as vectors in  $\mathbb{C}^n$ ,  $n = N_1 N_2$ , where  $\mathbb{C}^n$  stands for the space of complex-valued

vectors of length  $n$ . These vectors are obtained from the standard matrix representation for images by concatenating the columns of these matrices. We use bold lower case characters for these vectors. Thus,  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are vectorial representations for  $u_1[k]$  and  $u_2[k]$ .

The forward propagation of the wave field from the plane with the distribution  $\mathbf{u}_1$  gives  $\mathbf{u}_2$  as

$$\mathbf{u}_2 = \mathbf{A}_d \mathbf{u}_1, \quad (1)$$

where  $\mathbf{A}_d$  is a generic notation for a discrete forward propagation operator (complex-valued  $n \times n$  matrix,  $\mathbf{A}_d \in \mathbb{C}^{n \times n}$ ).

Eq.(1) corresponds to the convolution of the object distribution  $\mathbf{u}_0$  with the diffraction kernel of the wave field propagation operator  $\mathbf{A}_d$ . Depending on the used discretization of the Rayleigh-Sommerfield integral the operator  $\mathbf{A}_d$  in Eq.(1) can be: convolutional; angular spectrum decomposition (ASD); discrete diffraction transform given in the Fourier domain (F-DDT) [23], etc. We use the model (1) for both the reference and signal arms of the GI setup.

Let a complex-valued transmittance of SLM be given by a matrix  $M_r \in \mathbb{C}^{N_1 \times N_2}$ . Then for an uniform laser beam of the intensity 1 the wave field  $\mathbf{u}_r$  is defined by the equation  $\mathbf{u}_r = \mathbf{A}_d \mathbf{M}_r$ , where the vector  $\mathbf{M}_r \in \mathbb{C}^n$  stands for

the vectorized representation of the matrix  $M_r$ .

Let a set of  $K$  experiments be performed with different  $M_r$ ,  $r = 1, \dots, K$ .

The corresponding intensities at the output of the bucket sensor are of the form

$$o_r = \sum_{k=1}^n |\mathbf{u}_o[k]|^2 \cdot |\mathbf{u}_r[k]|^2 = \mathbf{b}_r^T \mathbf{c}, \quad r = 1, \dots, K. \quad (2)$$

where  $|\mathbf{u}_r[k]|^2$  are intensities of the speckle wave fields illuminating the object.

Here we use the vectorial notation  $\mathbf{c} = |\mathbf{u}_o|^2 \in \mathbb{R}^n$  for the squared magnitude of the transmission mask object, and  $\mathbf{b}_r = |\mathbf{u}_r|^2 \in \mathbb{R}^n$ . The intensities  $o_r = \mathbf{b}_r^T \mathbf{c}$  can be treated as projections of the squared modulus  $\mathbf{c} = |\mathbf{u}_o|^2$  on the random directions given by the intensity speckle vectors  $|\mathbf{u}_r|^2$ .

The cross-correlation of the observations  $o_r$  with the corresponding speckle wave fields  $\mathbf{u}_r$  is a fundamental idea for reconstruction of the object transparency  $|\mathbf{u}_o|^2$ . The analytical expression commonly used for this reconstruction in the form (e.g. [10])

$$|\hat{\mathbf{u}}_o[k]|^2 = \frac{1}{K} \sum_{r=1}^K (o_r - \bar{o}) |\mathbf{u}_r[k]|^2. \quad (3)$$

Here  $\bar{o} = \frac{1}{K} \sum_{r=1}^K o_r$  is the mean value of the observations. This reconstruction is calculated as a correlation function between the centered measurements and the intensities of the reference speckle fields  $|\mathbf{u}_r[k]|^2$ .

It can be seen that this estimate is biased and improperly scaled with respect to  $\mathbf{u}_o$ . A corrected version of (3) is discussed in Subsection 4.C.

*2.B. Poissonian and approximative Gaussian observations*

The Poissonian observations (2) take random integer values  $\tilde{o}_r$  interpreted as a counted number of photons detected by the sensor. This discrete distribution has a single parameter  $\lambda \geq 0$  and defined by the formula  $p(\tilde{o}_r = k) = \exp(-\lambda) \frac{\lambda^k}{k!}$ . Here  $p(\tilde{o}_r = k)$  is the probability that a random  $\tilde{o}_r$  takes value  $k$ ,  $k \geq 0$  is an integer. The parameter  $\lambda$  is the intensity flow of Poissonian random events.

For different experiments the parameter  $\lambda$  takes values  $\lambda_r = \mathbf{b}_r^T \mathbf{c}$ , and the probabilistic Poissonian observation model is given by the formula  $p(\tilde{o}_r = k) = \exp(-\mathbf{b}_r^T \mathbf{c} \chi) \frac{(\mathbf{b}_r^T \mathbf{c} \chi)^k}{k!}$ ,  $r = 1, \dots, K$ .

According to the properties of the Poissonian distribution, we have for the mean value and the variance of the observed  $\tilde{o}_r$ ,  $E\{\tilde{o}_r\} = var\{\tilde{o}_r\} = \mathbf{b}_r^T \mathbf{c} \chi$ . In these formulas  $\chi > 0$  is a scaling parameter. A larger  $\chi$  means a larger exposure time, and a larger number of the photons recorded by the sensor.

The probabilistic observation model corresponding to (2) can be written in

the form

$$\tilde{o}_r = \mathcal{Poisson}\{\mathbf{b}_r^T \mathbf{c} \chi\}, \quad (4)$$

where  $\mathcal{Poisson}\{\cdot\}$  stands for a generator of random Poissonian numbers with the corresponding intensity. Practically, instead of (4) the normalized intensity (intensity/sec) is measured as

$$o_r = \mathcal{Poisson}\{\mathbf{b}_r^T \mathbf{c} \chi\} / \chi. \quad (5)$$

It is not difficult to realize that  $E\{o_r\} = \mathbf{b}_r^T \mathbf{c}$  and  $var\{o_r\} = \mathbf{b}_r^T \mathbf{c} / \chi$ . Here  $E\{\cdot\}$  stands for operation of mathematical expectation. Thus, larger  $\chi$  (larger observation time) results in a smaller variance, and  $o_r$  is an unbiased estimate of  $\mathbf{b}_r^T \mathbf{c}$ . For the signal-to-noise ratio (SNR) we have  $SNR = E\{o_r\} / \sqrt{var\{o_r\}} = \sqrt{\mathbf{b}_r^T \mathbf{c} \chi}$ . Thus, smaller  $\chi$  means a smaller  $SNR$  and a relatively larger level of the noise. Note that the distribution of  $o_r$  defined in Eq.(5) is not Poissonian anymore, in particular because  $E\{o_r\} \neq var\{o_r\}$ .

For a sufficiently large  $\mathbf{b}_r^T \mathbf{c} \chi$ ,  $\mathbf{b}_r^T \mathbf{c} \chi > 1000$  (and even for much smaller  $\mathbf{b}_r^T \mathbf{c} \chi$ ), the Poissonian distribution is approximated with a good accuracy by the Gaussian distribution [24]. Then,  $o_r$  in Eq.(5) is also Gaussian with the distributions  $\mathcal{N}(\mathbf{b}_r^T \mathbf{c}, \mathbf{b}_r^T \mathbf{c} / \chi)$ .

The observed random  $o_r$  can be represented in the form with an additive

random noise

$$o_r = \mathbf{b}_r^T \mathbf{c} + \sigma_r \varepsilon_r, \quad r = 1, \dots, K, \quad (6)$$

$$\sigma_r^2 = \mathbf{b}_r^T \mathbf{c} / \chi. \quad (7)$$

where  $\varepsilon_r$  is the standard zero-mean Gaussian white random noise,  $\varepsilon_r \sim \mathcal{N}(0, 1)$ .

It was noticed in simulation experiments that the variances  $\sigma_r^2$  are nearly invariant and can be replaced by a constant value,  $\sigma_r^2 \simeq \sigma^2$ . It happens because for large  $n$  and under quite general assumptions  $\mathbf{b}_r^T \mathbf{c} \simeq \mathbf{E}\{\mathbf{b}_r^T \mathbf{c}\} = \mathbf{E}\{\mathbf{b}_r^T\} \mathbf{c} = \mathbf{E}\{\mathbf{b}^T\} \mathbf{c}$ , where  $\mathbf{b} = \mathbf{E}\{\mathbf{b}_r\}$ ,  $r = 1, \dots, K$ . Thus,  $\sigma_r^2$  can be taking asymptotically invariant,  $\sigma_r^2 \rightarrow_{n \rightarrow \infty} \mathbf{b}^T \mathbf{c} = \sigma^2$ .

Then, the model (6) is simplified to the form with an invariant standard deviation noise

$$o_r = \mathbf{b}_r^T \mathbf{c} + \sigma \varepsilon_r, \quad r = 1, \dots, K, \quad (8)$$

$$\sigma^2 = \mathbf{b}^T \mathbf{c} / \chi. \quad (9)$$

For a large  $\mathbf{b}_r^T \mathbf{c}$ ,  $o_r$  in Eq.(8) is nonnegative as it is in (4).

A good and unbiased estimate of  $\sigma^2$  from  $o_r$  is of the form

$$\hat{\sigma}^2 = \frac{1}{K} \sum_{r=1}^K o_r / \chi. \quad (10)$$

Restoration of the image  $\mathbf{c}$  from linear noisy observations of  $o_r$  is an inverse problems, and the linear observation operator is singular for  $K < n$ . We characterize *compressive sampling ratio* by the parameter  $\mu = n/K$  defined as a ratio of the image size  $n$  to the number of observations  $K$  used for reconstruction.

A thermal Gaussian random noise typical for digital sensors can be included in the models (6) and (8) as a Gaussian random summand.

### 2.C. Sparse object modeling

According to the sparse overcomplete approximation techniques the vector object  $\mathbf{c}$  can be represented in the following *synthesis* and *analysis* forms, respectively, [17, 18]:

$$\mathbf{c} = \mathbf{\Psi}\boldsymbol{\theta}, \tag{11}$$

$$\boldsymbol{\theta} = \mathbf{\Phi}\mathbf{c}. \tag{12}$$

Here,  $\mathbf{c} \in \mathbb{R}^n$ ,  $\boldsymbol{\theta} \in \mathbb{R}^m$ , and  $\mathbf{\Psi}$  and  $\mathbf{\Phi}$  are transform matrices of the corresponding sizes,  $n \times m$  and  $m \times n$ , respectively. The vector  $\boldsymbol{\theta}$ , usually called *spectrum*, gives the parameters for the parametric approximation of the image  $\mathbf{c}$  as  $\mathbf{c} = \mathbf{\Psi}\boldsymbol{\theta}$ . Thus,  $\mathbf{c} = \sum_{j=1}^m \mathbf{\Psi}_j \cdot \boldsymbol{\theta}_j$ , where  $\mathbf{\Psi}_j$  are the columns of the matrix  $\mathbf{\Psi}$ , and  $\boldsymbol{\theta}_j$  are the items of the vector  $\boldsymbol{\theta}$ .

The synthesis defines the image provided that the spectrum is given,  $\mathbf{c} = \mathbf{\Psi}\boldsymbol{\theta}$ . Contrary to it, the analysis defines the spectrum corresponding to a given image as  $\boldsymbol{\theta} = \mathbf{\Phi}\mathbf{c}$ .

The complexity of the image approximation  $\mathbf{c} = \sum_{j=1}^m \mathbf{\Psi}_j \cdot \boldsymbol{\theta}_j$  is characterized by the  $l_0$ -norm of the vector  $\boldsymbol{\theta}$ , denoted by  $\|\boldsymbol{\theta}\|_0$ , and calculated as a number of nonzero elements of the vector. The image model is named *over-complete* if  $m > n$  and *sparse* if the number of the nonzero (active) spectral elements of  $\boldsymbol{\theta}$  is smaller than the image size,  $\|\boldsymbol{\theta}\|_0 < n$ .

The  $l_1$ -norm of  $\boldsymbol{\theta}$  is defined as the sum of the absolute values of all items of the vector,  $\|\boldsymbol{\theta}\|_1 = \sum_s |\boldsymbol{\theta}_s|$ . Both these  $l_0$ - and  $l_1$ -norms are used in order to characterize *sparsity* of approximation. A smaller value of the norm means a higher sparsity of approximation (e.g. [25]).

In Eqs.(11)-(12) the transform matrices  $\mathbf{\Psi}$  and  $\mathbf{\Phi}$  define the bases (dictionary) for object approximations. In the classical approaches the orthonormal bases are the standard tools, where  $m = n$  and the analysis and synthesis transforms are such that  $\mathbf{\Phi}\mathbf{\Phi}^T = I_{n \times n}$ ,  $\mathbf{\Psi} = \mathbf{\Phi}^T$ . It is obvious that when we are looking for the sparsest approximation a larger set of potentially good basic function gives a better opportunity for model selection. It is recognized

that *overcomplete* representations for  $\mathbf{c}$  with  $m \gg n$  and linearly dependent  $\Psi_j$  form a much more powerful tool for advanced imaging than the classical orthonormal bases with  $m = n$ .

The concept of *frame* is a generalization of the classical bases especially developed for overcomplete (synthesis and analysis) representations with linearly dependent approximating functions (e.g. [26]).

There are special links between the analysis and synthesis frames. In particular, the requirement,  $\Psi \cdot \Phi = \mathbf{I}_{n \times n}$ , where  $\mathbf{I}_{n \times n}$  is the  $n \times n$  identity matrix, enables a perfect reconstruction of any  $\mathbf{Y}$  from the corresponding spectrum  $\boldsymbol{\theta}$ , indeed  $\mathbf{c} = \Psi\boldsymbol{\theta} = \Psi \cdot \Phi\mathbf{c} = \mathbf{c}$ . For details and applications of overcomplete, in particular, frame based modeling for imaging we refer to the recent book [25].

In our algorithm implementations we use for the analysis and synthesis the BM3D frames [17].

### 3. Variational object reconstruction

Different algorithms for reconstruction of the object transmittance  $\mathbf{c} = |\mathbf{u}_o|^2$  can be developed using the Poissonian (5) and approximative Gaussian models (6)-(7), (8)-(9). A comparative study shows that the algorithm based on

the Gaussian modeling with the invariant standard deviation demonstrates a very good performance. Computationally, this algorithm is much simpler than those for the Poissonian modeling and the Gaussian modeling with the varying standard deviations.

In this section we are focussed on the Gaussian modeling with the invariant standard deviation. The algorithms for the Gaussian modeling with the varying standard deviations and Poissonian observations are presented in Appendices A and B.

Let us start from the conventional CS technique based on the overcomplete modeling of the object (e.g. [10] and [11]). For the Gaussian noise distribution in Eq.(8)-(9) the maximum likelihood approach leads to the following objective function

$$J_G = L(\{o_r\}_1^K, \mathbf{c}) + \tau \cdot \|\Phi \mathbf{c}\|_{l_p}, \quad (13)$$

$$\sigma^2 = \mathbf{b}^T \mathbf{c} / \chi, \quad (14)$$

where  $\|\cdot\|_{l_p}$  stands for the  $l_p$ -norm,  $p = 0$  or  $p = 1$ , and  $\tau > 0$  is a regularization parameter.

The first summand in Eq.(13) is the *minus log-likelihood function* calculated using the joint probability density of the independent Gaussian observations

(8)-(9),  $L(\{o_r\}_1^K, \mathbf{c}) = -\ln \prod_{r=1}^K \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-1}{2\sigma^2}(o_r - \mathbf{b}_r^T \mathbf{c})^2)$ . The second summand in Eq.(13) is the penalty (regularization) term defining the complexity (sparsity) of the model for squared magnitude of the object transmittance  $\mathbf{c}$ .

The standard CS formulation of the reconstruction problem is of the form

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} J_G, \text{ or } \hat{\mathbf{c}} = \arg \min_{\mathbf{c} \geq \mathbf{0}} J_G. \quad (15)$$

The latter formulation is used if nonnegativity of  $\mathbf{c}$  is taken into consideration.

In this paper for derivation of the GI algorithm we use the original CS technique based on the multi-objective optimization which is very different from the single-objective optimization (15) conventional for CS. The advantage of this novel CS technique is demonstrated in the papers [17] and [18].

Let us introduce the following two objective functions:

$$\mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}) = L(\{\mathbf{o}_r\}_1^K, \mathbf{c}) + \frac{1}{2\gamma_0} \|\mathbf{c} - \Psi \boldsymbol{\theta}\|_2^2, \sigma^2 = \mathbf{b}^T \mathbf{c} / \chi, \quad (16)$$

$$\mathcal{L}_2(\mathbf{c}, \boldsymbol{\theta}) = \tau \cdot \|\boldsymbol{\theta}\|_{l_p} + \frac{1}{2} \|\boldsymbol{\theta} - \Phi \mathbf{c}\|_2^2. \quad (17)$$

Here  $\|\cdot\|_2^2$  stands for the squared Euclidean norm of the vector.

The first summand in Eq.(16) is the fidelity term from Eq.(13). The second summand corresponds to the transform modeling the constraint (11) linking  $\mathbf{c}$  with its spectrum  $\boldsymbol{\theta}$ . This link is given using the quadratic pe-

nalization for Eq.(11), where  $\gamma_0$  is a parameter of this penalization. Recall, that Eq.(11) gives the synthesis constraint. The quadratic penalization of constraint-equations is a common practice in order to reduce constrained optimization to unconstrained one. Smaller values of the corresponding penalization parameter  $\gamma_0$  mean that in the solution the corresponding equations will be fulfilled more accurately (e.g. [27]).

The first summand in (17) defines the complexity (sparsity) of the analysis model for the object squared magnitude  $\mathbf{c}$ . The second summand is the analysis constraint (12) again given using the quadratic penalization for Eq.(12).

The multi-objective optimization as it is proposed in [17] and [18] means the alternating minimization of the two objective functions  $\mathcal{L}_1$  and  $\mathcal{L}_2$  leading to the following iterative algorithm:

$$\mathbf{c}^{t+1} = \arg \min_{\mathbf{c} \geq 0} \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}^t), \quad (18)$$

$$\boldsymbol{\theta}^{t+1} = \arg \min_{\boldsymbol{\theta}} \mathcal{L}_2(\mathbf{c}^{t+1}, \boldsymbol{\theta}). \quad (19)$$

In Eq.(18) minimization of  $\mathcal{L}_1$  on  $\mathbf{c}$  inverses the image formation operator. Minimization of  $\mathcal{L}_2$  on  $\boldsymbol{\theta}$  filters the squared magnitude of the object transmittance function. Instead of minimization of a single-objective function as in Eq.(15) the algorithm (18)-(19) is looking for a quite different solution which

is a *fixed-point*  $(\boldsymbol{\theta}^*, \mathbf{c}^*)$  defined by the equations:

$$\mathbf{c}^* = \arg \min_{\mathbf{c} \geq \mathbf{0}} \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}^*), \quad (20)$$

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \mathcal{L}_2(\mathbf{c}^*, \boldsymbol{\theta}).$$

This fixed-point balances two different intentions: minimization of the fidelity term (accuracy of observation fitting) in  $\mathcal{L}_1$  and minimization of the complexity of the object model (defined in  $\mathcal{L}_2$  as  $\tau \cdot \|\boldsymbol{\theta}\|_{l_p}$ ), provided the analysis and synthesis restrictions.

In order to simplify the problem (18) we replace it by unconstrained optimization with projection of the solution on the quadrant of non-negative values. Instead of (18)-(19) we will consider the iterative procedure

$$\mathbf{c}^{t+1/2} = \arg \min_{\mathbf{c}} \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}^t), \quad \mathbf{c}^{t+1} = P_+\{\mathbf{c}^{t+1/2}\}, \quad (21)$$

$$\boldsymbol{\theta}^{t+1} = \arg \min_{\boldsymbol{\theta}} \mathcal{L}_2(\mathbf{c}^{t+1}, \boldsymbol{\theta}), \quad (22)$$

where  $P_+\{\mathbf{c}\}$  is an elementwise projector of the vector  $\mathbf{c}$ ,  $P_+\{\mathbf{c}[k]\} = \mathbf{c}[k]$  if  $\mathbf{c}[k] > 0$  and  $P_+\{\mathbf{c}[k]\} = 0$  otherwise.

### 3.A. Solution of optimization problems

Consider the optimization problems (21) and (22). Assume for a while that  $\sigma^2$  is known, e.g. calculated using the formula (10). Then

$$\begin{aligned} \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}) &= \frac{1}{2\sigma^2} \sum_{r=1}^K (o_r - \mathbf{c}^T \mathbf{b}_r)^2 + \\ &\frac{1}{2\gamma_0} \|\mathbf{c} - \Psi \boldsymbol{\theta}\|_2^2, \end{aligned} \quad (23)$$

and the minimum condition  $\partial \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}) / \partial \mathbf{c} = \mathbf{0}$  gives the linear equation

$$-\frac{1}{\sigma^2} \sum_{r=1}^K (o_r - \mathbf{c}^T \mathbf{b}_r) \mathbf{b}_r + \frac{1}{\gamma_0} (\mathbf{c} - \Psi \boldsymbol{\theta}) = \mathbf{0}. \quad (24)$$

Let us rewrite this equation as

$$\mathbf{c} = \Psi \boldsymbol{\theta} + \gamma_0 / \sigma^2 \sum_{r=1}^K (o_r - (\mathbf{c}^T \mathbf{b}_r)) \mathbf{b}_r \quad (25)$$

and multiply this equation by  $\mathbf{b}_k^T$  for  $k = 1, \dots, K$ . Introducing the variables  $x_r = \mathbf{c}^T \mathbf{b}_r$ , we arrive to the set of the  $K$  equations

$$\begin{aligned} \sum_{r=1}^K \left( \frac{1}{\sigma^2} \mathbf{b}_k^T \mathbf{b}_r + \frac{\delta_{k,r}}{\gamma_0} \right) x_r &= \sum_{r=1}^K \frac{1}{\sigma^2} o_r \mathbf{b}_k^T \mathbf{b}_r + \frac{1}{\gamma_0} \mathbf{b}_k^T \Psi \boldsymbol{\theta}, \\ k &= 1, \dots, K. \end{aligned}$$

Let  $\mathbf{B}$  be a  $n \times K$  matrix,  $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_K)$ , then the above equations can be written in the compact form

$$\left( \frac{1}{\sigma^2} \mathbf{B}^T \mathbf{B} + \frac{1}{\gamma_0} \mathbf{I}_{K \times K} \right) \cdot \mathbf{x} = \frac{1}{\sigma^2} \mathbf{B}^T \mathbf{B} \cdot \mathbf{o} + \frac{1}{\gamma_0} \mathbf{B}^T \Psi \boldsymbol{\theta}, \quad (26)$$

where  $\mathbf{x} = (x_1, \dots, x_K)^T$ .

With a given variable  $\mathbf{x}$  Eq.(25) is of the form

$$\mathbf{c} = \Psi\boldsymbol{\theta} + \frac{\gamma_0}{\sigma^2}\mathbf{B}(\mathbf{o} - \mathbf{x}). \quad (27)$$

Thus, the solution of Eq.(24) can be obtained in two sequential stages.

First,  $\mathbf{x} \in \mathbb{R}^K$  is calculated from (26), second the vector  $\mathbf{c} \in \mathbb{R}^n$  is calculated from (27). In Eq.(26) we need to invert  $K \times K$  matrix  $(\frac{1}{\sigma^2}\mathbf{B}^T\mathbf{B} + \frac{1}{\gamma_0}\mathbf{I}_{K \times K})$  instead of inversion of much larger  $n \times n$  matrix  $\sum_{r=1}^K(\frac{1}{\sigma^2}\mathbf{b}_r\mathbf{b}_r^T + \frac{1}{\gamma_0}\mathbf{I}_{n \times n})$  for Eq.(24). It defines the essential advantage of the derived two stage solution (26)-(27) in comparison with a direct solution of the original Eq.(24).

The problem (22) has the well known solution specified for  $l_0$  and  $l_1$  norms as the elementwise operator (e.g. [25])

$$\boldsymbol{\theta} = \mathfrak{Th}_\tau(\Phi\mathbf{c}) = \begin{cases} \mathfrak{Th}_\tau^{soft}(\Phi\mathbf{c}) = \text{sign}(\Phi\mathbf{c}) \circ \max(|\Phi\mathbf{c}| - \tau, 0), & \text{if } l_p = l_1, \\ \mathfrak{Th}_{\sqrt{2\tau}}^{hard}(\Phi\mathbf{c}) = \Phi\mathbf{c} \circ 1(|\Phi\mathbf{c}| \geq \sqrt{2\tau}), & \text{if } l_p = l_0, \end{cases} \quad (28)$$

where the indexes '*soft*' and '*hard*' indicate the type of the solution as the *soft*- or *hard-thresholding*. The threshold parameters for the soft- and hard thresholdings are different. In (28) ' $\circ$ ' stands for the elementwise multiplication of the elements of two vectors, and  $|\Phi\mathbf{c}|$  is a vector of the absolute values

of the elements of the vector  $\Phi\mathbf{c}$ .

### *3.B. Developed algorithm*

The successive steps of the proposed iterative algorithm are based on the above solutions for (21) and (22). In our implementation of the algorithm the analysis and synthesis operations, thresholding and design of the analysis and synthesis frames (matrixes  $\Phi$  and  $\Psi$ ) are integrated in a single block, which we call *BM3D-filter*. Then the developed algorithm can be presented in the

following compact form:

**Algorithm: CSGI<sub>inv</sub>**

**Input:**  $\mathbf{o}$ ,  $\mathbf{B}$ ,  $\mathbf{c}^{init}$ ,  $\sigma_{init}^2$

**Set**  $t = 0$ ,  $\mathbf{v}_0^0 = \mathbf{c}^{init}$ ,  $(\sigma^0)^2 = \sigma_{init}^2$

**Repeat until convergence:**

$$1. \left( \frac{1}{(\sigma^t)^2} \mathbf{B}^T \mathbf{B} + \frac{1}{\gamma_0} \mathbf{I}_{K \times K} \right) \mathbf{x}^{t+1} = \frac{1}{(\sigma^t)^2} \mathbf{B}^T \mathbf{B} \mathbf{o} + \frac{1}{\gamma_0} \mathbf{B}^T \mathbf{v}_0^t;$$

$$2. \mathbf{c}^{t+1/2} = \mathbf{v}_0^t + \frac{\gamma_0}{(\sigma^t)^2} \mathbf{B}(\mathbf{o} - \mathbf{x}^{t+1}),$$

$$\mathbf{c}^{t+1} = P_+ \{ \mathbf{c}^{t+1/2} \};$$

$$3. \mathbf{v}_0^{t+1} = \mathbf{BM3D-filter}(\mathbf{c}^{t+1});$$

$$4. (\sigma^{t+1})^2 = \mathbf{b}^T \mathbf{c}^{t+1} / \chi;$$

$$t = t + 1.$$

The input variables of the algorithm are: the intensity observation vector  $\mathbf{o} \in \mathbb{R}_+^K$ , the speckle wave field intensities  $\mathbf{B} \in \mathbb{R}^{K \times n}$ , calculated by propagation of the SLM phase mask radiations  $M_r$  in the scenario of Fig.2 or measured in the scenario of Fig.1, the initial guess  $\mathbf{c}^{init} \in \mathbb{R}_+^n$  for  $|\mathbf{u}_0|^2$ , and the estimate of the noise variance  $\sigma_{init}^2$  calculated according to Eq.(10).

Steps 1 and 2 corresponds to the solution (26)-(27) of the problem (21) with minimization of  $\mathcal{L}_1$  with respect to  $\mathbf{c}$ .

At Step 3 the analysis  $\Phi$  and synthesis  $\Psi$  transforms are constructed by BM3D-filter for the image (input of this filter)  $\mathbf{c}^{t+1}$ . The BM3D analysis operation defines the spectrum  $\Phi\mathbf{c}^{t+1}$ , which after thresholding gives the filtered spectrum  $\boldsymbol{\theta}^{t+1} = \mathfrak{T}_{h_\tau}(\Phi\mathbf{c}^{t+1})$ . The BM3D synthesis operations  $\mathbf{v}_0^{t+1} = \Psi\boldsymbol{\theta}^{t+1}$  defines the output of BM3D-filter.

In Step 4 the noise variance  $(\sigma^{t+1})^2$  is updated according to the current estimate of  $\mathbf{c}$ .

The initial guess  $\mathbf{c}^{init}$  is calculated according to  $t_0$  steps of a *no-filtering* option of the CS $\text{GI}_{\text{inv}}$  algorithm. This no-filtering option is a simplified version of CS $\text{GI}_{\text{inv}}$  obtained by dropping the filtering step 3 and assuming that

$$\mathbf{v}^{t+1} = \mathbf{c}^{t+1}.$$

BM3D-filter allows two different modes: first, with the fixed analysis and synthesis frames and, second, with the design of the analysis and synthesis frames.

The mode with the design is compulsory for the iteration  $t = t_0 + 1$ . The BM3D frames are data dependent and for this step they are obtained for

the initial guess  $\mathbf{c}^{init}$ . These obtained frames could be fixed up to the end of reconstruction.

The frames also can be updated by BM3D-filter after a number of iterations. Then these new frames depend on the reconstructions  $\mathbf{c}^t$ , where  $t$  is a number of iteration where the update of the frames is produced.

It was noticed that only a few updates should be used because a larger number of updates does not improve the reconstruction and even can results in their degradation. Note also that the frame design takes time and slows reconstruction process.

We use the name  $\mathbf{CSGI}_{inv}$  for the proposed algorithm as the abbreviation for **C**ompressive **S**ensing **G**host **I**maging with invariant standard deviation.

The reconstruction algorithm  $\mathbf{CSGI}_{var}$  for Gaussian observations with varying standard deviation and the algorithm  $\mathbf{CSGI}_{poiss}$  for Poissonian observations are presented in Appendices A and B, respectively.

#### 4. Simulation experiments

In the presented experimental results we use two types of transmission mask objects: binary  $26 \times 44$  test-image *TUT* (abbreviation for Tampere University of Technology) and a  $64 \times 64$  fragment of the gray-scale test-image *Camera-*

man . In our vectorial notation they form vectors  $\mathbf{c} \in \mathbb{R}_+^n$  with  $n = 26 \times 44$  and  $n = 64 \times 64$ , respectively.

Simulation is addressed to the single-arm scenario (Fig.2). The pixelation of modeling is defined by pixels of SLM. We assume that these pixels are square  $\Delta_x \times \Delta_x$ ,  $\Delta_x = 8 \mu m$ , with 100% fill factors. The wavelength  $\lambda = .532 \mu m$  corresponds to a green laser. “*In-focus*” distance for the considered lensless setup is calculated as  $d_f = N \cdot \Delta_x^2 / \lambda$ . In particular,  $d_f = 7.7 mm$  for  $N = 64$ .

The F-DDT technique is applied for modeling of the free space forward propagation (operator  $A_d$  in Eq.(1)). Recall that this technique enables the exact wave field propagation for pixelated sensor and object wave fields [23].

The phase modulation masks  $M_r$ ,  $r = 1, \dots, K$ , are generated according to the formula

$$M_r[k_1, k_2] = \exp(j2\pi\varphi_r[k_1, k_2]),$$

$$1 \leq k_1 \leq N_1, 1 \leq k_2 \leq N_2,$$

where the random phase  $\varphi_r[k_1, k_2]$  is uniformly distributed over the interval  $[0, 1)$  and independent for all  $k_1, k_2$  and  $r$ .

The random masks  $M_r$ ,  $r = 1, \dots, K$ , are generated only once and fixed through all image reconstruction experiments. It is done in order to enable

the reproducible and accurately comparable results for different parameters of the experimental setup. In particular, it is used for optimization of the free propagation distance  $d$ . The observations  $o_r$  are Poissonian generated according to Eq.(5). The number of experiments/observations  $K$  is defined as  $K = n/\mu$ , where the compressive sampling ratio  $\mu = 1, 2, 4, 8, 16$  for *TUT* and  $\mu = 2, 4, 8, 16$  for *Cameraman* test-images.

We consider the distance  $d$  between the SLM and the sensor in Fig. 2 as a design parameter and produced experiments for  $d = d_f \cdot k_d$ ,  $k_d = 1, 2, 3, 5, 10$ .  $k_d = 1$  corresponds to *in-focus* imaging, while  $k_d > 1$  gives *out-of-focus* imaging.

Parameter  $\chi$  defines the level of the randomness in the observations. In our experiments  $\chi = 10^2, 10^4, 10^6$ . For test-image *TUT* it corresponds to SNR equal about 15.6, 35.6 and 55.6 dB. Thus, for  $\chi = 10^2$  we obtain the most noisy observations, while  $\chi = 10^6$  corresponds to the nearly noiseless case. Note that the noise level obtained even for  $\chi = 10^4$  is quite essential for the considered ill-posed (singular) inverse imaging problems.

In what follows we show the results obtained for the hard-shresholding filtering only corresponding to use the  $l_0$ -norm in (28). Experimental study

shows that the  $l_0$ -norm usually results in better image reconstructions than those for the  $l_1$ -norm.

The computational complexity of the algorithm is characterized by the time required for 100 iterations, including the initialization iterations. For an image  $64 \times 64$  it takes about 200 sec. provided  $\mu = 2$  ( $K = 2028$ ) and about 60 sec. provided  $\mu = 4$  ( $K = 1024$ ) for the computer and the software used in the experiments: Intel Core 2Duo E8400 @ 3GHz, RAM 4GB, Windows Xp SP3, Matlab 7.9.0 (R2010b). The computationally most intensive parts of the algorithm (BM3D-frame design, analysis and synthesis operations) have been written in C++.

Two parameters are used in order to tune the algorithm: weight  $1/\gamma_0$  and threshold  $\tau$ . Overall, larger values of the threshold parameter  $\tau$  and smaller values of  $\gamma_0$  result in a stronger smoothing of reconstruction. For selection of the parameters  $\gamma_0$  and  $\tau$  we use the *oracle rule* assuming that the true image is given and the corresponding PSNR is maximized. In this case the shown results demonstrate the potential of the algorithm. Practically, these parameters can be selected by special tests looking for parameter values optimal for various classes of objects.

We make our MATLAB programs for the demo version of our algorithms publicly available for testing: <http://www.cs.tut.fi/~lasip/DDT/>.

#### 4.A. Performance of **CSGI**<sub>inv</sub> algorithm

We use *peak signal-to-noise ratio* (PSNR) as a criterion for the accuracy of the object reconstruction. PSNR is calculated as a ratio between the maximum possible power of the signal and the power of reconstruction error. PSNR is expressed in the logarithmic decibel scale,  $PSNR = 20 \log_{10} \max_k |u_o[k]|^2 / RMSE$  dB, where  $\max_k |u_o[k]|^2$  is the maximum value of the squared magnitude of the object, and  $RMSE$  is the root-mean-squared error of  $|u_o|^2$  reconstruction.

Usually the values of  $PSNR$  about  $25 \div 30$  dB are evidence of a good quality reconstruction, and 40 dB and higher values correspond to a nearly perfect reconstruction.

In Tables 1-3 we show PSNR values obtained for *TUT* test-image with  $\chi = 10^2, 10^4, 10^6$ , respectively. The rows of the tables correspond to distances  $d = d_f \cdot k_d$ ,  $k_d = 1, 3, 5, 10$ , and the columns show results for different compressive sampling ratio  $\mu = 1, 2, 4, 8, 16$ .

The cells of these tables where  $PSNR \leq 20$  dB are marked by *fail*, because

of a low quality imaging in these cases.

For dependence of PSNR with respect to the free propagation distance  $d$ , we can note that there exists an optimal distance in each column giving the maximum PSNR value. Provided fixed other parameters,  $d$  varies the size of speckles in  $u_r$ . In general, larger speckles appear for larger  $d$ . The existence of the optimal value for  $d$  is intuitively obvious.

The probabilistic modeling of speckles presented in Goodman's book [28] allows to calculate speckle size, speckle correlation, speckle resolution, etc. However, it is quite difficult to apply these results for evaluation of the real GI accuracy for a particular image. The developed simulation instruments can be successfully used for this sort of problems, and Tables 1-3 demonstrates it. The found optimal  $d$  define speckle sizes which are the best for a particular image and particular experimental parameters.

It is interesting to note that the best results can be achieved for compressive sampling ratio  $\mu$  which is not compulsory equal to 1 (the case when the number of observations is equal to image size). In particular, the best  $PSNR = 45.82$  dB for  $\chi = 10^2$  is obtained for  $d = 3d_f$  and  $\mu = 2$ .

The overall best results for  $\chi = 10^2, 10^4, 10^6$  are, respectively  $PSNR =$

31.60, 45.82, and 58.85 dB. Naturally, lower levels of noise allow to get better results.

The visual quality of imaging is illustrated in Fig.3-Fig.5. For noiseless data with  $\chi = 10^6$  we obtain the perfect reconstruction even for the compressive sampling ratio  $\mu = 4$ , in the same scenario but with the compressive sampling ratio  $\mu = 8$ , the result shown in Fig.4 is good but not perfect with  $PSNR = 29.3$  dB.

Imaging of gray-scale objects is a more difficult problem as compared with imaging of a binary object such as *TUT*. In Fig.5 we show the reconstruction for the *Cameraman* test-image. The visual quality is good with quite high value of PSNR. This result is shown for noisy data with  $\chi = 10^4$  and the compressive sampling ratio  $\mu = 2$ .

Provided the fixed phase masks  $M_r$ ,  $r = 1, \dots, K$ , the only source of randomness in observations and image reconstructions are defined by Poissonian observations calculated according to Eq.(5). We checked by Monte Carlo experiments that  $PSNR$  values are nearly identical for different trails of the random observations with a difference in  $PSNR$  of about  $1\% \div 2\%$ .

Another interesting point concerns the compressibility of the overcomplete

image modeling when the used dictionaries (frames  $\Phi$  and  $\Psi$ ) are truly redundant.

Let  $\|\theta\|_{l_0}/M$  be *compressive spectral ratio*, where  $M$  is the total size of the spectrum  $\theta$  and  $\|\theta\|_{l_0}$  is a number of the nonzero elements of the  $\theta$  obtained after the thresholding. This *compressive spectral ratio* shows what a portion of the spectra is actually used for approximation. Remind that the image reconstruction is sparse  $\|\theta\|_{l_0} < n$ , where  $n$  is the image size.

In our experiments we obtain the following figures which are typical for good quality maging. For TUT test-image with  $\mu = 4$ :  $\|\theta\|_{l_0}/M \simeq 10$ ,  $\|\theta\|_{l_0} \simeq 260$ ,  $M \simeq 2600$  and  $PSNR \simeq 39$  dB. This reconstruction is sparse because  $\|\theta\|_{l_0} \simeq 250 < n = 1144$  and shows a good value for the compressive spectral ratio equal to 10.

For Cameraman test-image with  $\mu = 2$  :  $\|\theta\|_{l_0}/M \simeq 5$ ,  $\|\theta\|_{l_0} \simeq 1300$ ,  $M \simeq 6500$  and  $PSNR \simeq 30$  dB. This reconstruction is sparse because  $\|\theta\|_{l_0} \simeq 1300 < n = 6400$  and also shows a good value for the compressive spectral ratio equal to 5.

Thus, BM3D frames enable a good compressibility both in the image and the spectral domains, a good sparsity of reconstruction as well as a good

quality imaging.

#### 4.B. $\mathbf{CSGI}_{inv}$ versus $\mathbf{CSGI}_{var}$ and $\mathbf{CSGI}_{poiss}$

Comparison of the algorithms  $\mathbf{CSGI}_{inv}$  and  $\mathbf{CSGI}_{var}$  shows that the advantage which could be gained by  $\mathbf{CSGI}_{var}$ , based on the more accurate observation modeling, is quite negligible provided that the level of the noise is not high,  $\chi > 10^3 \div 10^4$ . The PSNR values can be improved by few percents only.

A more valuable PSNR improvement by  $\mathbf{CSGI}_{var}$  can be achieved for noisier data,  $\chi < 10^3$ . However, the visual quality of this imaging is poor. We wish to note that the computational complexity of  $\mathbf{CSGI}_{var}$  is essentially higher than the complexity of  $\mathbf{CSGI}_{inv}$ , because the matrix to be inverted  $(\mathbf{B}^T \mathbf{B} \mathbf{D}^t + \frac{1}{\gamma_0} \mathbf{I}_{K \times K})$  is different for each iteration while the corresponding matrix  $(\mathbf{B}^T \mathbf{B} + \frac{1}{\gamma_0} \mathbf{I}_{K \times K})$  in  $\mathbf{CSGI}_{inv}$  is invariant and can be inverted only once.

The complexity of  $\mathbf{CSGI}_{poiss}$  is even higher than the complexity of  $\mathbf{CSGI}_{var}$  because a set of the quadratic (not linear as it is for Gaussian approximations) equation (43) should be solved for every iteration.

The PSNR advantage by  $\mathbf{CSGI}_{poiss}$  versus  $\mathbf{CSGI}_{inv}$  could be demonstrated

only for quite small  $\chi$ , however, despite of this numerical advantage, the visual quality of the imaging is poor.

Our conclusion is that the exposure time in experiments should be large enough in order to enable a lower level of the randomness in observations and the possibility of good imaging. Then the algorithm  $\mathbf{CSGI}_{inv}$  is able to give good results and application of the algorithm  $\mathbf{CSGI}_{var}$  and  $\mathbf{CSGI}_{poiss}$  has no sense.

#### 4.C. Cross-correlation reconstructions

Let us rewrite the cross-correlation GI algorithm (3) as  $\frac{1}{K} \sum_{r=1}^K (o_r - \bar{o}_K) \mathbf{b}_r$ , where  $\mathbf{b}_r = |\mathbf{u}_r|^2 \in \mathbb{R}^n$ , and modify this algorithm to the following form

$$|\hat{\mathbf{u}}_o[k]|^2 = \frac{1}{K} \sum_{r=1}^K (o_r - \bar{o}_K) (\mathbf{b}_r - \bar{\mathbf{b}}_K), \quad (29)$$

$$\bar{\mathbf{b}}_K = \frac{1}{K} \sum_{r=1}^K \mathbf{b}_r, \quad \bar{o}_K = \frac{1}{K} \sum_{r=1}^K o_r \quad (30)$$

Simple manipulations show that  $\frac{1}{K} \sum_{r=1}^K (o_r - \bar{o}_K) (\mathbf{b}_r - \bar{\mathbf{b}}_K) = \frac{1}{K} \sum_{r=1}^K (\mathbf{b}_r - \bar{\mathbf{b}}_K)^T \mathbf{c} (\mathbf{b}_r - \bar{\mathbf{b}}_K)$ ,  $\mathbf{c} = |\mathbf{u}_o|^2$ . Then the expectation of (29) is calculated as

$$E\{|\hat{\mathbf{u}}_o|^2\} = \mathbf{D}_K \mathbf{c}, \quad \mathbf{D}_K = E\left\{ \frac{1}{K} \sum_{r=1}^K (\mathbf{b}_r - \bar{\mathbf{b}}_K) (\mathbf{b}_r - \bar{\mathbf{b}}_K)^T \right\}.$$

Using this result we propose a modified cross-correlation algorithm in the

form

$$|\hat{\mathbf{u}}_o[k]|^2 = \hat{\mathbf{D}}_K^+ \cdot \frac{1}{K} \sum_{r=1}^K (o_r - \bar{o}_K)(\mathbf{b}_r - \bar{\mathbf{b}}_K), \quad (31)$$

$$\hat{\mathbf{D}}_K = \frac{1}{K} \sum_{r=1}^K (\mathbf{b}_r - \bar{\mathbf{b}}_K)(\mathbf{b}_r - \bar{\mathbf{b}}_K)^T.$$

Here  $\hat{\mathbf{D}}_K$  is an asymptotically unbiased estimate of  $\mathbf{D}_K$ , and  $\hat{\mathbf{D}}_K^+$  is a pseudoinverse (or regularized inverse) of  $\mathbf{D}_K$ . Then, the reconstruction (31) is asymptotically unbiased and properly scaled with respect to  $|\mathbf{u}_o|^2$ . Thus, we modify the estimate (3) in two ways. First,  $\mathbf{b}_r$  is replaced by a centered random  $\mathbf{b}_r - \bar{\mathbf{b}}_K$ , and second, the  $n \times n$  matrix  $\hat{\mathbf{D}}_K^+$  makes the reconstruction asymptotically unbiased.

It can be proved that the estimate  $|\hat{\mathbf{u}}_o|^2$  converges to  $|\mathbf{u}_o|^2$  in the mean squared sense, i.e.  $E\{|||\hat{\mathbf{u}}_o|^2 - |\mathbf{u}_o|^2||_2^2\} \rightarrow 0$  as  $K \rightarrow \infty$ . However, the experiments show that the convergence rate with respect to  $K$  is very slow. A number of experiments required for a good accuracy should be in orders larger than the image size. In Fig.6 and Fig.7 we show the performance of the algorithm (31) for two test images *TUT* and *Cameraman*. The imaging achieved even for  $K = 30000$  is quite poor visually and numerically in PSNR values. This number of experiments is much larger than the sizes of these test images,  $K \gg n$ .

Comparison with reconstructions obtained in Subsection 4.A show that the developed  $\text{CSGI}_{inv}$  algorithm yields incomparably better results provided much smaller number of experiments  $K$ .

## 5. Conclusion

The computational GI using a phase spatial light modulator (SLM) for a wave field modulation is considered. Three maximum likelihood algorithms are developed, respectively for Poissonian observation modeling and approximate Gaussian ones with varying and invariant standard deviations. An original compressive imaging technique developed in this paper results in the iterative algorithms decoupling inverse of the image formation operator and filtering. A sparse and overcomplete object modeling based on BM3D frames enables the advanced high accuracy reconstruction and sharp imaging even for quite large values of compressive sampling ratio. It is demonstrate that a Gaussian approximation for Poissonian observations using an invariant variance results in the efficient and comparative simple in implementation algorithm.

## Appendix A. Algorithm for Gaussian observation modeling with varying standard deviations

For the varying standard deviations the Gaussian minus log-likelihood is calculated as  $L(\{\mathbf{o}_r\}_1^K, \mathbf{c}) \simeq \sum_{r=1}^K [\frac{1}{2\sigma_r^2}(o_r - \mathbf{c}^T \mathbf{b}_r)^2 + \ln \sigma_r]$ ,  $\sigma_r^2 = \mathbf{b}_r^T \mathbf{c} / \chi$ .

The unconstrained minimum condition  $\partial \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}) / \partial \mathbf{c} = \mathbf{0}$  in the problem (21) gives the equation nonlinear with respect to  $\mathbf{c}$  :

$$-\sum_{r=1}^K \frac{1}{\sigma_r^2} (o_r - \mathbf{c}^T \mathbf{b}_r) \mathbf{b}_r + \sum_{r=1}^K \left[ \frac{-1}{\sigma_r^2} (o_r - \mathbf{c}^T \mathbf{b}_r)^2 + 1 \right] \left( \frac{1}{2\mathbf{b}_r^T \mathbf{c}} \mathbf{b}_r \right) + \frac{1}{\gamma_0} (\mathbf{c} - \Psi \boldsymbol{\theta}) = 0. \quad (32)$$

It can be shown that  $(o_r - \mathbf{c}^T \mathbf{b}_r)^2$  is of the order  $\sigma_r^2$  and  $(o_r - \mathbf{c}^T \mathbf{b}_r)$  is of the order  $\sigma_r$ . Then for a large  $\chi$  the first summand in Eq.(32) is of the order  $\theta(\sqrt{\chi})$  and the second summand is of the order  $\theta(1)$ . It follows that for large  $\chi$  the second summand in Eq.(32) can be omitted, and the equation becomes linear with respect to  $\mathbf{c}$  provided given  $\sigma_r$ :

$$-\sum_{r=1}^K \frac{1}{\sigma_r^2} (o_r - \mathbf{c}^T \mathbf{b}_r) \mathbf{b}_r + \frac{1}{\gamma_0} (\mathbf{c} - \Psi \boldsymbol{\theta}) = 0. \quad (33)$$

After manipulations similar to the ones used in Subsection 3.A the solution

of Eq.(33) can be given by the two stage calculations

$$\begin{aligned}
 (\mathbf{B}^T \mathbf{B} \mathbf{D} + \frac{1}{\gamma_0} \mathbf{I}_{K \times K}) \cdot \mathbf{x} &= \mathbf{B}^T \mathbf{B} \mathbf{D} \cdot \mathbf{o} + \frac{1}{\gamma_0} \mathbf{B}^T \Psi \boldsymbol{\theta}, \\
 \mathbf{D} &= \text{diag}\{\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_K^2}\},
 \end{aligned} \tag{34}$$

and

$$\mathbf{c} = \Psi \boldsymbol{\theta} + \gamma_0 \mathbf{B} \mathbf{D} (\mathbf{o} - \mathbf{x}). \tag{35}$$

First,  $\mathbf{x} \in \mathbb{R}^K$  is calculated from (34), and, second, the vector  $\mathbf{c} \in \mathbb{R}^n$  is calculated using (35). If  $\sigma_r = \sigma$ ,  $r = 1, \dots, K$ , then  $\mathbf{D} = \mathbf{I}_{K \times K} / \sigma^2$  and Eqs.(34)-(35) are identical to Eqs.(26)-(27).

The solution of the problem (22) has the form (28). The GI reconstruction

algorithm is given in the form

**Algorithm: CSGI<sub>var</sub>**

**Input :**  $\mathbf{o}$ ,  $\mathbf{B}$ ,  $\mathbf{c}^{init}$ ,

**Set**  $t = 0$ ,  $\mathbf{c}^0 = \mathbf{v}_0^0 = \mathbf{c}^{init}$ ,  $(\sigma_r^0)^2 = \mathbf{b}_r^T \mathbf{c}^{init} / \chi$

**Repeat until convergence:**

$$1. (\mathbf{B}^T \mathbf{B} \mathbf{D}^t + \frac{1}{\gamma_0} \mathbf{I}_{K \times K}) \cdot \mathbf{x}^{t+1} =$$

$$\mathbf{B}^T \mathbf{B} \mathbf{D}^t \cdot \mathbf{o} + \frac{1}{\gamma_0} \mathbf{B}^T \mathbf{v}_0^t;$$

$$2. \mathbf{c}^{t+1/2} = \mathbf{v}_0^t + \gamma_0 \mathbf{B} \mathbf{D}^t (\mathbf{o} - \mathbf{x}^t),$$

$$\mathbf{c}^{t+1} = P_+ \{ \mathbf{c}^{t+1/2} \};$$

$$3. \mathbf{v}_0^{t+1} = \mathbf{BM3D-filter}(\mathbf{c}^{t+1});$$

$$4. (\sigma_r^{t+1})^2 = \mathbf{b}_r^T \mathbf{c}^{t+1} / \chi,$$

$$\mathbf{D}^{t+1} = \text{diag} \left\{ \frac{1}{(\sigma_1^{t+1})^2}, \dots, \frac{1}{(\sigma_K^{t+1})^2} \right\};$$

$$t = t + 1.$$

The algorithm is initiated by  $\mathbf{c}^{init}$  obtained by the CSGI<sub>inv</sub> algorithm. Steps 1 and 2 of the algorithm are different from the corresponding steps of the CSGI<sub>inv</sub> algorithm by the varying weights  $\mathbf{D}^t$  depending on the variances  $(\sigma_r^t)^2$ .

Step 3 is identical to the corresponding step 3 in the  $\text{CSGI}_{\text{inv}}$  algorithm. In Step 4 the variances  $(\sigma_r^t)^2$  are recalculated accordingly to the updated values of the reconstruction  $\mathbf{c}^t$ . Meaning and using of BM3D-filter is discussed in subsection 3.B.

We use the name  $\text{CSGI}_{\text{var}}$  for the developed algorithm, where "var" emphasizes that the varying variances in  $\mathbf{D}^{t+1}$  is the main difference between the algorithms  $\text{CSGI}_{\text{inv}}$  and  $\text{CSGI}_{\text{var}}$ .

## Appendix B. Algorithm for Poissonian observation modeling

The minus loglikelihood for the Poissonian observations Eq.(4) has the form

$$L(\{\mathbf{o}_r\}_1^K, \mathbf{c}) = \sum_{r=1}^K (-\tilde{o}_r \log \tilde{\mathbf{b}}_r^T \mathbf{a} + \tilde{\mathbf{b}}_r^T \mathbf{a}), \quad (36)$$

where  $\tilde{\mathbf{b}}_r = \mathbf{b}_r \chi$ .

In the framework of the multi-objective optimization formalism, for the Poissonian observations and sparse modeling of the object we introduce two objective functions

$$\mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}) = \sum_{r=1}^K (-\tilde{o}_r \log \tilde{\mathbf{b}}_r^T \mathbf{c} + \tilde{\mathbf{b}}_r^T \mathbf{c}) + \frac{1}{2\tilde{\gamma}_0} \|\mathbf{c} - \Psi\boldsymbol{\theta}\|_2^2, \quad (37)$$

and

$$\mathcal{L}_2(\mathbf{c}, \boldsymbol{\theta}) = \tau \cdot \|\boldsymbol{\theta}\|_0 + \frac{1}{2} \|\boldsymbol{\theta} - \Phi\mathbf{c}\|_2^2. \quad (38)$$

The first summand in Eq.(37) corresponds to the Poissonian fidelity term (36). The second summand is the transform modeling constraint (11) linking the squared magnitude  $\mathbf{c}$  with its spectrum  $\boldsymbol{\theta}$ . This link is given using the quadratic penalization function, where  $\tilde{\gamma}_0$  is a parameter of this penalization.

Let us consider minimization of  $\mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta})$ . The necessary minimum condition for (37) has the form

$$\partial\mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta})/\partial\mathbf{c} = \sum_{r=1}^K \left(-\tilde{o}_r \frac{1}{\tilde{\mathbf{b}}_r^T \mathbf{c}} \tilde{\mathbf{b}}_r + \tilde{\mathbf{b}}_r\right) + \frac{1}{\tilde{\gamma}_0} (\mathbf{c} - \Psi\boldsymbol{\theta}) = 0. \quad (39)$$

From (39) we obtain the equations for  $\tilde{\mathbf{b}}_k^T \mathbf{c}$

$$\begin{aligned} & \tilde{\mathbf{b}}_k^T \partial \mathcal{L}_1(\mathbf{c}, \boldsymbol{\theta}) / \partial \mathbf{c} = \\ & \sum_{r=1}^K (1 - \tilde{o}_r \frac{1}{\tilde{\mathbf{b}}_r^T \mathbf{c}}) \tilde{\mathbf{b}}_k^T \tilde{\mathbf{b}}_r + \frac{1}{\tilde{\gamma}_0} (\tilde{\mathbf{b}}_k^T \mathbf{c} - \tilde{\mathbf{b}}_k^T \Psi \boldsymbol{\theta}) = 0, \\ & k = 1, \dots, K. \end{aligned} \tag{40}$$

With notations  $\tilde{\mathbf{b}}_k^T \mathbf{c} = \tilde{x}_k$ ,  $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_K)^T$ , and using a  $n \times K$  matrix,  $\tilde{\mathbf{B}} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_K)$ , this set of the equation can be rewritten in the compact form

$$\tilde{\mathbf{B}}^T \tilde{\mathbf{B}}(\tilde{\mathbf{o}}./\tilde{\mathbf{x}}) - \frac{\mathbf{1}}{\tilde{\gamma}_0} \tilde{\mathbf{x}} = \tilde{\mathbf{B}}^T \tilde{\mathbf{B}} \mathbf{1}_{K \times 1} - \frac{1}{\tilde{\gamma}_0} \tilde{\mathbf{B}}^T \Psi \boldsymbol{\theta}, \tag{41}$$

where  $\mathbf{1}_{K \times 1}$  is a  $K \times 1$  vector of 1.

Here  $\tilde{\mathbf{o}}./\tilde{\mathbf{x}}$  stands for the elementwise division of the corresponding items of two vectors.

If  $x_k$  are found then it follows from (39) that  $\mathbf{c} = \Psi \boldsymbol{\theta} + \tilde{\gamma}_0 \sum_{r=1}^K (o_r \frac{1}{\tilde{x}_r} - 1) \tilde{\mathbf{b}}_r$  or in the matrix form

$$\mathbf{c} = \Psi \boldsymbol{\theta} + \tilde{\gamma}_0 (\tilde{\mathbf{B}}(\tilde{\mathbf{o}}./\tilde{\mathbf{x}}) - \tilde{\mathbf{B}} \mathbf{1}_{K \times 1}). \tag{42}$$

Now let us make the following replacement of the variables

$$\tilde{\mathbf{B}} = \mathbf{B} \chi, \quad \tilde{\mathbf{x}} = \mathbf{x} \chi, \quad \tilde{\mathbf{b}} = \mathbf{b}_r \chi, \quad \tilde{\mathbf{o}} = \mathbf{o} \chi, \quad \gamma_0 = \tilde{\gamma}_0 \chi.$$

Then the equations (41)-(42) take the final form

$$\mathbf{B}^T \mathbf{B}(\mathbf{o}/\mathbf{x}) - \frac{1}{\gamma_0} \mathbf{x} = \mathbf{B}^T \mathbf{B} \mathbf{1}_{K \times 1} - \frac{1}{\gamma_0} \mathbf{B}^T \Psi \boldsymbol{\theta}. \quad (43)$$

$$\mathbf{c} = \Psi \boldsymbol{\theta} + \gamma_0 (\mathbf{B} \cdot (\mathbf{o}/\mathbf{x}) - \mathbf{B} \mathbf{1}_{K \times 1}), \quad (44)$$

where  $\mathbf{o} = \tilde{\mathbf{o}}/\chi$  as it is in Eq.(5).

Thus, instead of the linear equations (26)-(27) and (34)-(35) derived for the Gaussian observations we obtain a set of the quadratic equations (43)-(44) for the Poissonian case. The solution of the problem (22) has the form (28).

The algorithm proposed for the Poissonian data can be as follows

**Algorithm: CSGI<sub>poiss</sub>**

**Input :**  $\mathbf{o}$ ,  $\mathbf{b}_r$ ,  $\mathbf{v}_0^0 = \mathbf{c}^{init}$

**Repeat until convergence:**

1.  $\mathbf{B}^T \mathbf{B}(\mathbf{o}/\mathbf{x}^{t+1}) - \frac{1}{\gamma_0} \mathbf{x}^{t+1} = \mathbf{B}^T \mathbf{B} \mathbf{1}_{K \times 1} - \frac{1}{\gamma_0} \mathbf{B}^T \mathbf{v}^t;$

2.  $\mathbf{c}^{t+1/2} = \mathbf{v}^t + \gamma_0 \chi^2 (\mathbf{B}(\mathbf{o}/\mathbf{x}^{t+1}) - \mathbf{B} \mathbf{1}_{K \times 1}),$

$$\mathbf{c}^{t+1} = P_+ \{\mathbf{c}^{t+1/2}\};$$

3.  $\mathbf{v}_0^{t+1} = \mathbf{BM3D}\text{-filter}(\mathbf{c}^{t+1});$

$$t = t + 1.$$

The initial guess  $\mathbf{c}^{init}$  is calculated using one of the Gaussian algorithms.

Steps 1 and 2 of the algorithm are obtained from the solutions (43)-(44).

Step 3 is identical to the corresponding step 3 in the  $\text{CSGI}_{\text{inv}}$  algorithm. We use the abbreviation  $\text{CSGI}_{\text{poiss}}$  for this algorithm, where *poiss* stands for Poissonian.

## A. Acknowledgments

This work was supported by the Academy of Finland: project no. 213462, 2006-2011 (Finnish Programme for Centres of Excellence in Research) and project no. 138207, 2011-2014.

## References

1. T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, "Optical imaging by means of two-photon quantum entanglement," *Phys. Rev. A* 52, R3429–R3432 (1995).
2. A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato, "Correlated imaging, quantum and classical," *Phys. Rev. A* 70, 013802 (2004).
3. Y. Cai and S.-Y. Zhu, "Ghost imaging with incoherent and partially coherent light radiation," *Phys. Rev. E* 71, 056607 (2005).
4. L. Basano and P. Ottonello, "A conceptual experiment on single-beam coincidence detection with pseudothermal light," *Opt. Express* 19, 12386–12394 (2009).
5. B. I. Erkmen and J. H. Shapiro, "Unified theory of ghost imaging with Gaussian-state light," *Phys. Rev. A* 77, 043809 (2008).
6. B. I. Erkmen and J. H. Shapiro, "Ghost imaging: from quantum to classical to computational," *Advances in Optics and Photonics*, vol. 2, no. 4, pp. 405-450 (2010).
7. Y. Shih, "The physics of ghost imaging," *Advances in Lasers and Electro Optics*, Nelson Costa and Adolfo Cartaxo (Ed.), ISBN: 978-953-307-088-9, Publisher: InTech, pp.549-594 (2010).
8. J. H. Shapiro, "Computational ghost imaging," *Phys. Rev. A* 78, 061802(R) (2008).

9. Y. Bromberg, O. Katz, and Y. Silberberg, "Ghost imaging with a single detector," *Phys. Rev. A* **79**, 053840 (2009).
10. O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Appl. Phys. Lett.* **95**, 113110 (2009).
11. L. Jiying, Z. Jubo, L. Chuan, and H. Shisheng, "High-quality quantum-imaging algorithm and experiment based on compressive sensing," *Optics Letters*, Vol. 35, No. 8, pp. 1206-1208 (2010).
12. E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
13. D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
14. S. Gazit, A. Szameit, Y. C. Eldar, M. Segev, "Super-resolution and reconstruction of sparse sub-wavelength images, " *Optics Express* **17**, pp. 23920-23946 (2009).
15. R. Willett, R. Marcia, and J. Nichols, "Compressed sensing for practical optical systems: a tutorial," *Optical Engineering*, vol. 50, no. 7, pp. 072601 1-13 (2011).
16. K. W. C. Chan, M. N. O'Sullivan, and R. W. Boyd, "Optimization of thermal ghost imaging: high-order correlations vs. background subtraction," *Optics Express*, vol. 18, no. 6, pp. 5562-5573 (2010).
17. A. Danielyan, Katkovnik, V. and Egiazarian, K., "BM3D frames and variational image deblurring," *IEEE Trans. Image Processing*, vol. 21, no. 4, pp. 1715-1728 (2012).
18. V. Katkovnik and J. Astola, "High-accuracy wave field reconstruction: decoupled inverse imaging with sparse modeling of phase and amplitude," *JOSA A* Vol. 29, Iss. 1, pp. 44-54

- (2012).
19. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095 (2007).
  20. M. Figueiredo and J. Bioucas-Dias, “Restoration of Poissonian images using alternating direction optimization”, *IEEE Trans. on Image Processing*, vol. 19, no. 12, 3133-3145, (2010).
  21. Z. Harmany, R. Marcia, and R. Willett, “This is SPIRAL-TAP: sparse Poisson intensity reconstruction algorithms – Theory and Practice,” *IEEE Trans. on Image Processing*, vol. 21, no.3, 1084-1096 (2012).
  22. M. Raginsky, R. Willett, Z. Harmany, and R. Marcia, “Compressed sensing performance bounds under Poisson noise,” *IEEE Trans. on Signal Processing*, vol. 58, no. 8, pp. 3990-4002 (2010).
  23. V. Katkovnik, J. Astola, and K. Egiazarian, “Discrete diffraction transform for propagation, reconstruction, and design of wave field distributions,” *Appl. Opt.* **47**, pp. 3481-3493 (2008).
  24. A. Papoulis, *Probability, random variables, and stochastic processes* (New York: McGraw-Hill, 1991).
  25. M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer (2010).
  26. D. Han, K. Kornelson, D. Larson, E. Weber, *Frames for undergraduates* (Student Mathematical Library, AMS, 2007).

27. D. P. Bertsekas, *Nonlinear programming* (Second ed., Cambridge, MA.: Athena Scientific, 1999).
28. J.W. Goodman, *Speckle phenomena in optics* (Ben Roberts, Greenwood Village, CO, 2007).

## List of Figure Captions

Fig.1 Two-arm SLM ghost-imaging setup.

Fig.2 Single-arm computational ghost-imaging setup.

Fig.3 TUT test-image,  $\mu = 4$ ,  $d = 5d_f$ ,  $\chi = 10^6$ . First row from left to right:  $\text{CSGI}_{inv}$  reconstruction and true image; second row from left to right: cross-sections of the true image (solid, red in colour) and the reconstruction (dash, blue in colour), and PSNR versus number of iterations.

Fig.4 TUT test-image,  $\mu = 8$ ,  $d = 5d_f$ ,  $\chi = 10^6$ . First row from left to right:  $\text{CSGI}_{inv}$  reconstruction and true image; second row from left to right: cross-sections of the true image (solid, red in colour) and the reconstruction (dash, blue in colour), and PSNR versus number of iterations.

Fig.5 Fragment of cameraman test-image,  $\mu = 2$ ,  $d = d_f$ ,  $\chi = 10^4$ . First row:  $\text{CSGI}_{inv}$  reconstruction and true image; second row: cross-sections of the true image (solid, red in colour) and the reconstruction (dash, blue in colour), and PSNR versus number of iterations.

Fig. 6 Cross-correlation reconstruction of the TUT test-image,  $d = 3d_f$ ,  $\chi = 10^6$ : reconstruction and PSNR versus the number of experiments  $K$ . In the cross-section the true image is shown by solid line (red in color) and the

reconstruction by dash line (blue in color).

Fig. 7 Cross-correlation reconstruction of the Cameraman test-image,  $d = d_f$ ,  $\chi = 10^4$ : reconstruction and PSNR versus the number of experiments  $K$ . In the cross-section the true image is shown by solid line (red in color) and the reconstruction by dash line (blue in color)..

## List of Table Captions

Table 1. Object transparency (TUT test image) reconstruction (PSNR values) for various distances of the free space wavefield propagation  $d$  and compressive sampling ratios  $\mu, \chi = 10^2$ .

Table 2. Object transparency (TUT test image) reconstruction (PSNR values) for various distances of the free space wavefield propagation  $d$  and compressive sampling ratios  $\mu, \chi = 10^4$ .

Table 3. Object transparency (TUT test image) reconstruction (PSNR values) for various distances of the free space wavefield propagation  $d$  compressive sampling ratios  $\mu, \chi = 10^6$ .

Table 1. Object transparency (TUT test image) reconstruction (PSNR values) for various distances of the free space wavefield propagation  $d$  and compression ratios  $\mu, \chi = 10^2$ .

$d \setminus \mu$	1	2	4	8	16
$d_f$	31.60	26.73	fail	fail	fail
$3d_f$	29.81	24.86	19.77	fail	fail
$5d_f$	23.80	22.21	fail	fail	fail
$10d_f$	fail	fail	fail	fail	fail

Table 2. Object transparency (TUT test image) reconstruction (PSNR values) for various distances of the free space wavefield propagation  $d$  and compressive sampling ratios  $\mu, \chi = 10^4$ .

$d \setminus \mu$	1	2	4	8	16
$d_f$	43.91	44.28	30.69	fail	fail
$3d_f$	45.53	45.82	38.94	19.05	fail
$5d_f$	36.61	34.19	27.55	22.82	fail
$10d_f$	20.23	20.41	19.95	fail	fail

Table 3. Object transparency (TUT test image) reconstruction (PSNR values) for various distances of the free space wavefield propagation  $d$  and compressive sampling ratios  $\mu$ ,  $\chi = 10^6$ .

$d \setminus \mu$	1	2	4	8	16
$d_f$	50.39	58.85	39.99	fail	fail
$3d_f$	50.42	52.36	42.76	20.92	fail
$5d_f$	52.47	42.53	52.66	29.65	fail
$10d_f$	21.35	21.10	21.08	20.95	fail

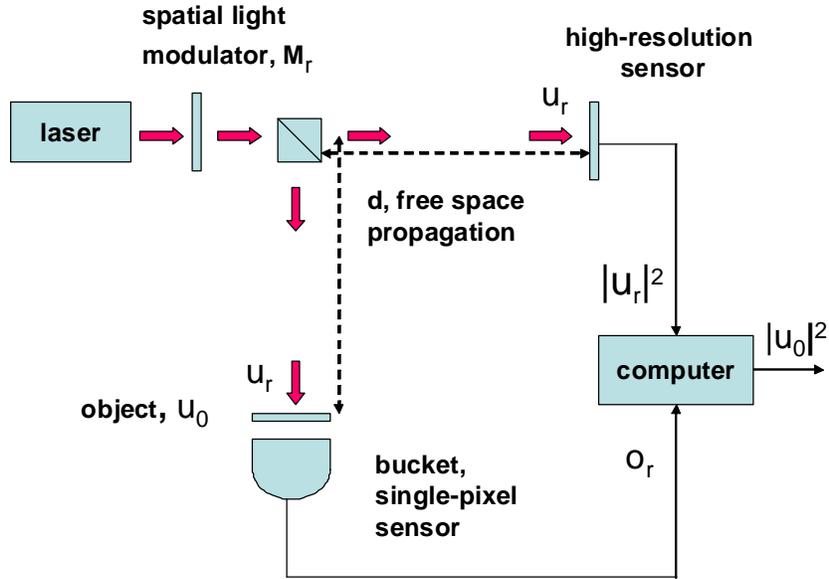


Fig. 1. Two-arm SLM ghost-imaging setup.

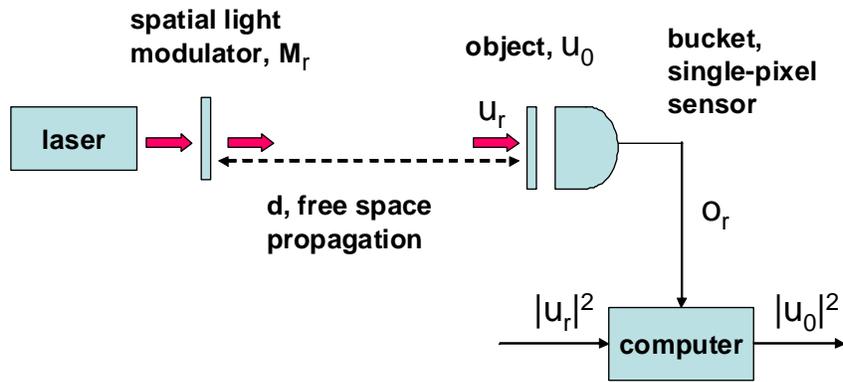


Fig. 2. Single-arm computational ghost-imaging setup.

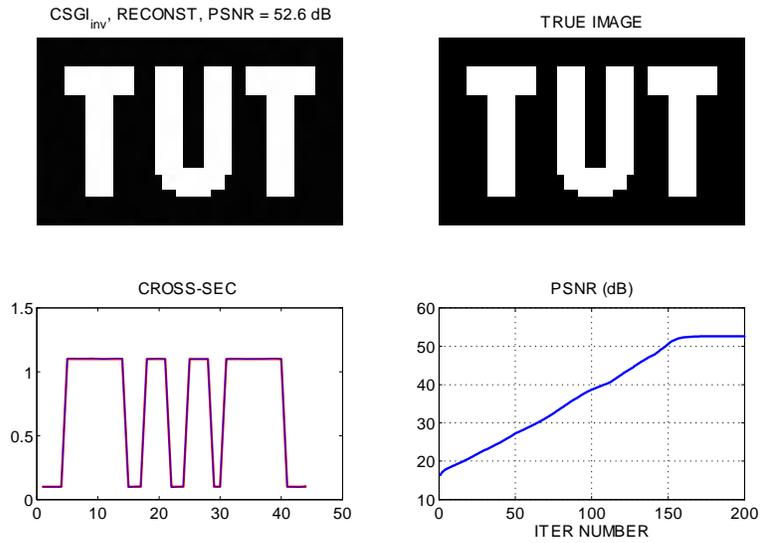


Fig. 3. TUT test-image,  $\mu = 4$ ,  $d = 5z_f$ ,  $\chi = 10^6$ . First row from left to right: CSGI<sub>inv</sub> reconstruction and true image; second row from left to right: cross-sections of the true image (solid, red in colour) and the reconstruction (dash, blue in colour), and PSNR versus number of iterations.

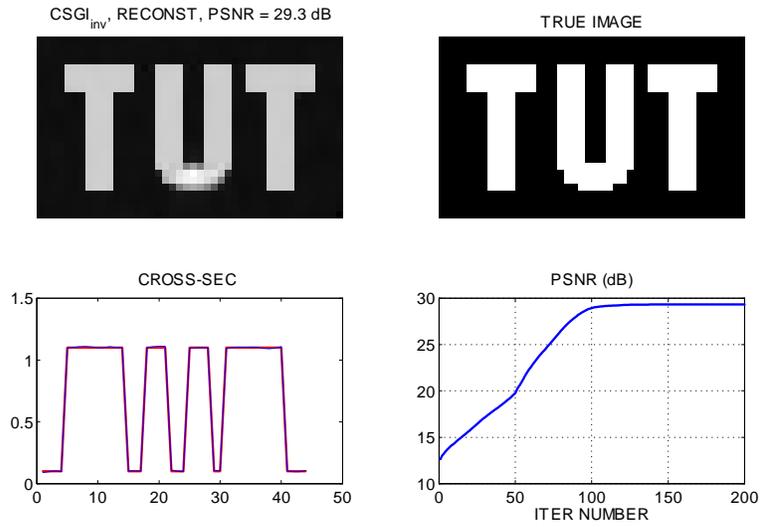


Fig. 4. TUT test-image,  $\mu = 8$ ,  $d = 5d_f$ ,  $\chi = 10^6$ . First row from left to right:  $\text{CSGI}_{inv}$  reconstruction and true image; second row from left to right: cross-sections of the true image (solid, red in colour) and the reconstruction (dash, blue in colour), and PSNR versus number of iterations.

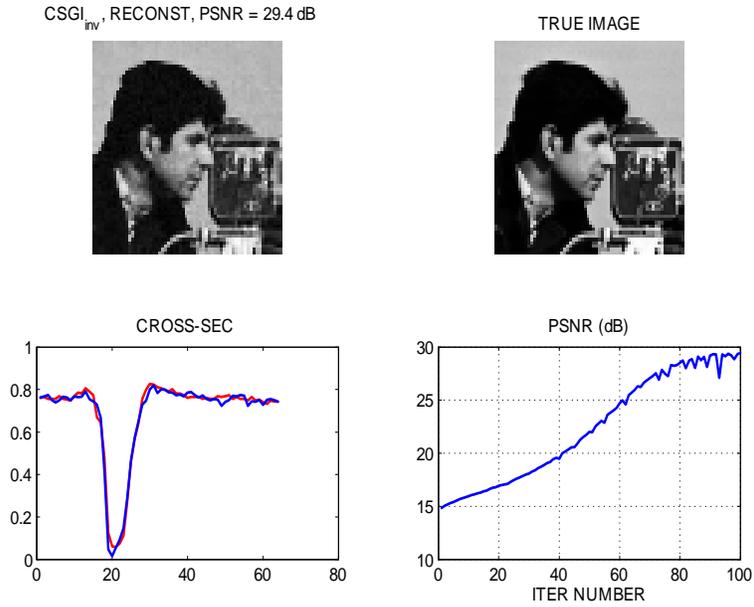


Fig. 5. Fragment of cameraman test-image,  $\mu = 2$ ,  $d = d_f$ ,  $\chi = 10^4$ . First row:  $\text{CSGI}_{inv}$  reconstruction and true image; second row: cross-sections of the true image (solid, red in colour) and the reconstruction (dash, blue in colour), and PSNR versus number of iterations.

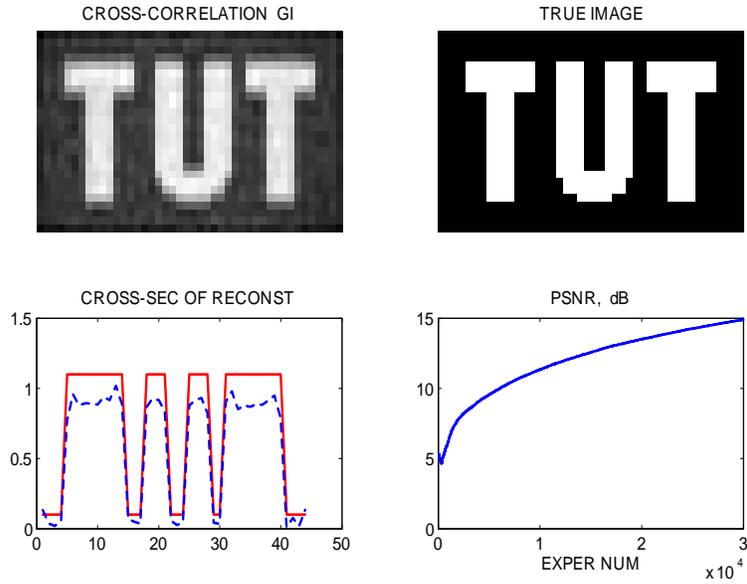


Fig. 6. Cross-correlation reconstruction of the TUT test-image,  $d = 3d_f$ ,  $\chi = 10^6$ : reconstruction and PSNR versus the number of experiments  $K$ . In the cross-section the true image is shown by solid line (red in color) and the reconstruction by dash line (blue in color).

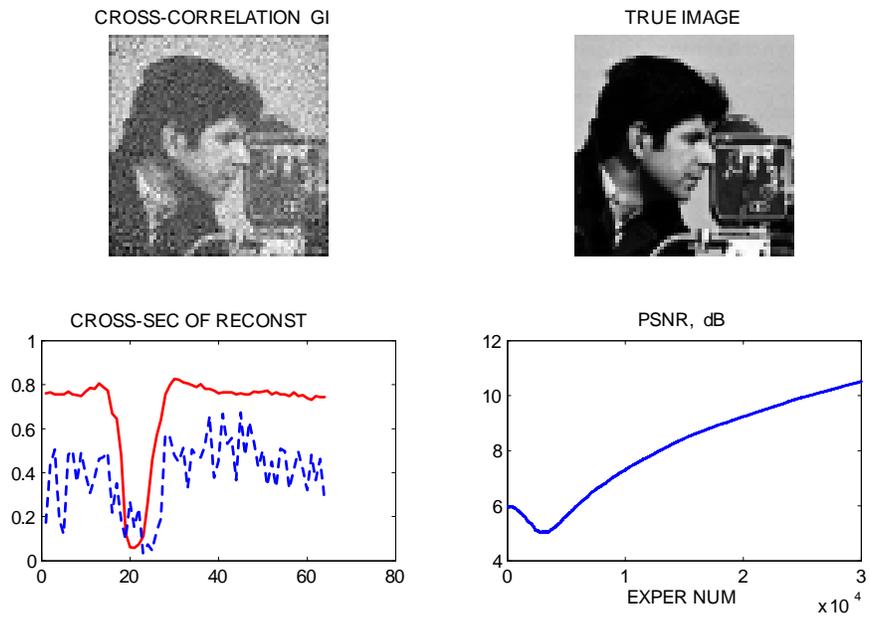


Fig. 7. Cross-correlation reconstruction of the Cameraman test-image,  $d = d_f$ ,  $\chi = 10^4$ : reconstruction and PSNR versus the number of experiments  $K$ . In the cross-section the true image is shown by solid line (red in color) and the reconstruction by dash line (blue in color).