

Received February 14, 2020, accepted February 24, 2020, date of publication February 27, 2020, date of current version March 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976870

# Dual Imaging—Can Virtual Be Better Than Real?

IVAN RALASIC<sup>ID</sup>, (Student Member, IEEE), MATEA DONLIC<sup>ID</sup>, (Student Member, IEEE),  
AND DAMIR SERSIC<sup>ID</sup>, (Member, IEEE)

Faculty of Electrical Engineering and Computing, University of Zagreb, 10000 Zagreb, Croatia

Corresponding author: Ivan Ralasic (ivan.ralasic@fer.hr)

This work was supported in part by the Croatian Science Foundation under Project IP-2019-04-6703, and in part by the European Regional Development Fund under Grant KK.01.1.1.01.0009 (DATACROSS).

**ABSTRACT** In this paper, a dual imaging technique is used for high-precision reconstruction of an observed 3D scene. In contrast to stereo vision, dual imaging systems use a camera and a projector instead of a camera pair. We propose a multiresolution approach based on the sum-to-one transform, coupled with compressive sensing principles, for efficient estimation of the light transport matrix (LTM). The LTM contains information on both optical systems and the 3D scene. In our setup, the camera sensor is intentionally chosen to be low resolution to prove the future use of inexpensive sensors in nonvisible regions of the light spectrum, as well as the potential design of simplified multiview and light field acquisition systems. We show that a high-precision estimation of the LTM from a reduced set of measurements is possible. Virtual measurements, instead of physical, are conducted to obtain the 3D reconstruction. We show that 3D scene reconstruction from the proposed virtual measurements corresponds with the actual physical acquisition. Moreover, this approach provides much more detail in the reconstruction. The computational complexity of the proposed methods is reduced to such a level that practical implementations are feasible.

**INDEX TERMS** 3D reconstruction, camera-projector system, compressive sensing, dual imaging, light transport.

## I. INTRODUCTION

The resolution of imaging sensors represents a limit on the amount of detail that we can reconstruct from an observed scene. This limit is particularly troublesome for imaging in nonvisible regions of the light spectrum and in multiview and light field imaging systems. In this paper, based on a novel LTM estimation technique, we introduce virtual measurements for 2D and 3D scene reconstruction to supplant real measurements. Hence, we aim to offer an affirmative answer to the following question: Can virtual be better than real?

Helmholtz reciprocity [1], [2] enables efficient modeling of light transport between a light source and a photosensor. It states that incoming and outgoing light paths can be considered as reversals of each other without affecting the bidirectional reflectance distribution function (BRDF). Helmholtz reciprocity is typically summarized by an equation describing the symmetry of radiance between incoming and outgoing

directions  $\omega_i$  and  $\omega_o$ , respectively:

$$f_r(\omega_i \rightarrow \omega_o) = f_r(\omega_o \rightarrow \omega_i), \quad (1)$$

where  $f_r$  represents the BRDF of the surface.

Dual photography [3] is a photographic technique that uses Helmholtz reciprocity to efficiently model light transport in camera-projector systems. This approach enables powerful postprocessing capabilities of the light transport in an observed scene, including virtual interchange of the camera and projector viewpoints and virtual relighting of the scene. The light transport matrix is a linear operator that describes geometric and photometric properties of the light propagation within the observed scene. LTM is a relatively large matrix whose size depends on the camera and projector resolutions.

Brute-force LTM acquisition is based on pixel scanning, where a single projector pixel is illuminated at a time and a camera is used to capture the individual responses. This approach presents several drawbacks, including long acquisition times, limited contrast in projection devices and limited sensitivity of the camera sensor. Several algorithms for LTM estimation have been proposed in the literature to address these issues. In their initial work, Sen *et al.* [3] proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Long Wang<sup>ID</sup>.

two improvements to the naïve LTM acquisition approach: fixed pattern scanning and adaptive multiplexed illumination. Fixed pattern scanning introduces parallelism into the acquisition process by illuminating multiple projector pixels at the same time. The basic assumption is that the individual projector pixels affect only a small, localized region of the scene from the camera viewpoint, which does not hold in the general case.

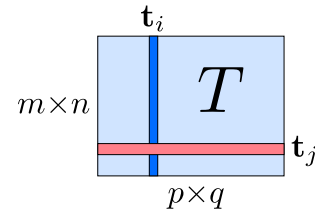
Adaptive multiplexed illumination performs online adaptation of the patterns to determine which projector pixels can be illuminated simultaneously. Adaptive patterns relax the assumption of direct illumination. One drawback of this approach is that individual camera images have to be processed during the acquisition to ensure separability of the light transport of each projector pixel.

Several papers discuss the application of compressive sensing (CS) principles in the LTM acquisition process [4]–[7]. The basic idea of CS LTM estimation is to perform a nonadaptive measurement process followed by an offline LTM reconstruction that estimates the light transport coefficients using sparsity and compressibility priors. Nonadaptive measurements simplify the implementation since individual frames captured by the camera do not need to be processed during acquisition. Peers *et al.* [5] illuminate the scene using patterns defined by various compression bases, while Sen and Darabi [4] decouple the choice of illumination patterns from the choice of compression basis. Decoupling the measurement and compression basis provides improved flexibility in the overall measurement process, overcoming quantization and dynamic range problems in the projection devices when projecting nonbinary patterns. Although the CS methods reduce the number of projections needed for efficient LTM estimation, the reconstruction involves iterative and computationally complex sparse optimization, which is time-consuming.

In [6], the authors introduce a homogeneous light transport matrix, in which the LTM and background projector light are described by a single matrix and estimated using CS principles. That approach enables efficient removal of the projector background light from the LTM estimate. Moreover, the authors exploit the LTM to obtain 3D measurements of the observed scene. In [7], the authors propose a multiscale approach to the LTM estimation that significantly accelerates the light transport acquisition. The basic idea of the multiscale approach is to project multiresolution measurement patterns and estimate the light transport for a pyramid of projector resolutions from lower to higher resolutions.

In our research, we are particularly interested in two goals: efficient modeling of the light transport in camera-projector systems and exploitation of geometric and photometric relations in the light transport for efficient 2D and 3D reconstruction that surpasses the capabilities of physical imaging sensors.

This paper is organized as follows: Section II offers an introduction to the dual imaging modality. We provide a short overview of CS LTM estimation, followed by our proposal



**FIGURE 1.** Column  $\mathbf{t}_i$  in  $\mathbf{T}$  corresponds to a vectorized image of the camera  $\mathbf{c}'_i$  when a single projector pixel in  $\mathbf{p}'_i$  is illuminated in primal configuration. Row  $\mathbf{t}_j$  in  $\mathbf{T}$  corresponds to a vectorized image of the projector  $\mathbf{p}''_j$  when a single camera pixel in  $\mathbf{c}''_j$  is “illuminated” in the dual configuration.

for fast LTM estimation using STOne transformation and compressive sensing. Section III covers basic principles of 3D imaging, namely, phase shifting profilometry and camera-projector calibration. To improve the quality of the 3D reconstruction in our camera-projector system, we propose 3D reconstruction in the dual world of the camera-projector system. The experimental results are presented in Section IV. Section V offers discussion related to the obtained results, while Section VI concludes the paper.

## II. DUAL IMAGING

We observe a system comprised of a camera with spatial resolution of  $m \times n$  pixels and a projector with resolution of  $p \times q$  pixels. LTM enables modeling of the light transport in the camera-projector system using a linear equation:

$$\mathbf{c}' = \mathbf{T}\mathbf{p}' \quad (2)$$

In (2), vector  $\mathbf{p}'$  (size  $pq \times 1$ ) denotes the vectorized projector image, vector  $\mathbf{c}'$  (size  $mn \times 1$ ) denotes the captured image on the camera sensor and matrix  $\mathbf{T}$  (size  $mn \times pq$ ) denotes the LTM containing the transport coefficients between every pair of camera and projector pixels at the resolution of both devices. Note that (2) describes the primal configuration of the camera-projector setup, which is denoted by the prime superscript ( $'$ ), following the notation from [3].

Camera and projector can be virtually interchanged following the Helmholtz reciprocity by transposing the light transport matrix:

$$\mathbf{p}'' = \mathbf{T}^\top \mathbf{c}'', \quad (3)$$

where  $\mathbf{T}^\top$  denotes the reciprocal light transport, while the dual configuration is denoted by the double prime superscript ( $''$ ). The structure of the light transport matrix  $\mathbf{T}$  is explained in Fig. 1.

### A. COMPRESSIVE SENSING ESTIMATION OF LIGHT TRANSPORT MATRIX

One existing challenge in dual imaging is efficient acquisition and reconstruction of the high-resolution light transport matrix. Instead of projecting a total number of  $pq$  measurement patterns defined by the number of projector pixels, CS methods for LTM estimation consider acquisition of  $k \ll (pq)$

images using different illumination patterns. This approach significantly reduces the number of projections required for efficient LTM estimation.

Similar to (2), we can write the light transport equation in matrix form as:

$$\mathbf{C} = \mathbf{TP}, \quad (4)$$

where  $\mathbf{C}$  is an  $mn \times k$  matrix whose columns represent the individual captured images, and  $\mathbf{P}$  is an  $pq \times k$  matrix whose columns represent the individual projected patterns. After transposing both sides ( $\mathbf{C}^\top = \mathbf{P}^\top \mathbf{T}^\top$ ), measurement for a single camera pixel over time can be rewritten as:

$$\mathbf{c}_i = \mathbf{P}^\top \mathbf{t}_i, \quad (5)$$

where  $\mathbf{t}_i$  is the  $pq \times 1$  reflectance function of the  $i^{\text{th}}$  camera pixel,  $\mathbf{c}_i$  is the  $k \times 1$  measurement vector, and  $1 \leq i \leq mn$ . In the following text,  $\Phi = \mathbf{P}^\top$  denotes the measurement matrix containing projected patterns with the standard notation used in the CS framework for brevity and congruence.

After capturing a set of measurements  $\mathbf{c}_i$ , sparse optimization is performed in order to obtain the LTM estimate. Notice that there are  $mn$  independent subproblems to solve, one for each camera pixel (5). The fact that these problems are independent offers opportunity for massive parallelization in the LTM reconstruction.

In this paper, we assume that the LTM is sparse in the spatial domain. The unconstrained form of the sparse optimization problem for light transport estimation can be written as:

$$\min_{\mathbf{t}_i} \|\mathbf{c}_i - \Phi \mathbf{t}_i\|_2^2 + \lambda \|\mathbf{t}_i\|_1, \quad i = 1, \dots, (mn) \quad (6)$$

where the  $l_1$ -norm is used as the regularizer since it is a convex relaxation of the sparsity inducing  $l_0$ -norm [8]. In other words, our goal is to find a solution that is sparse but congruent with the obtained measurements.

Sparse optimization algorithms are mostly iterative and computationally complex. For example, in [4], the authors report LTM reconstruction times on the order of three hours for a 24-node Linux cluster, with each node containing 2 Xeon 5140 CPUs running at 2.33 GHz for moderate camera and projector resolutions. Obviously, this fact limits the real-world applications of the dual photography.

A multiscale approach to the LTM estimation reduces the computational complexity. In [7], the authors propose to use measurement patterns with several increasing resolutions. Measurements obtained using the lower resolution patterns are used to mask zero coefficients in the LTM. In the subsequent resolutions, only parts of the LTM that were nonzero in the previous resolutions are considered for reconstruction. Although this procedure improves the reconstruction speed, it presents drawbacks in terms of accuracy. That is, the number of patterns projected in the highest resolution is often not sufficient to obtain an LTM estimation with acceptable accuracy.

In contrast, the approach proposed in this paper enables both fast and accurate estimation of the LTM. To achieve

fast LTM estimation, we utilize sum-to-one (STOne) transformation in the measurement design. Instead of projecting patterns at different resolutions, all measurement patterns used in our approach have the highest possible resolution achievable by the projector. Utilizing STOne measurement patterns enables us to implicitly obtain multiresolution LTM estimates. Ultimately, we are interested in exploiting the LTM for 3D reconstruction of the observed scene. We exploit dual imaging principles in order to obtain high-resolution 3D reconstructions that surpass the resolution of the imaging sensor.

## B. FAST LTM ESTIMATION USING STONE TRANSFORMATION

In [9], the authors propose a novel multiscale CS operator for images that enables twofold reconstruction. First, low-resolution previews of the image can be reconstructed at the Nyquist rate using the sum-to-one transform (STOne). Previews are fast conventional reconstructions that do not utilize sparse optimization while requiring only one measurement per pixel in the low-resolution preview. Alternately, previews can be enhanced using conventional CS methods in order to overcome the Nyquist limit. In [9], the authors demonstrate the efficiency of the STOne transform by constructing a real-time CS video camera. They use STOne transformation in order to reconstruct fast video previews of the observed scene at low resolution, which enables real-time analysis and better video coding.

STOne transform is a fast orthogonal transformation with the property that each row of the transform matrix has a unit sum. In [9], the authors construct the CS measurement matrix using the STOne transform. The basic construction block can be written as:

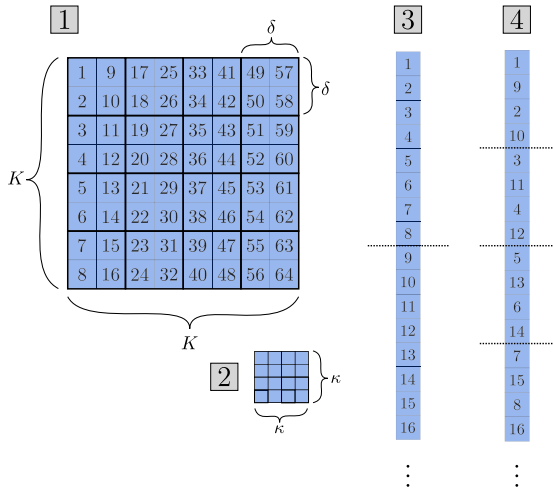
$$\mathbf{S}_4 = \frac{1}{2} \begin{bmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}, \quad (7)$$

and is used to construct STOne transform matrices at different resolutions using the Kronecker delta product  $\otimes$  as:

$$\mathbf{S}_{4^{k+1}} = \mathbf{S}_4 \otimes \mathbf{S}_{4^k}. \quad (8)$$

In the measurement matrix design, we closely follow the procedure described in [9]. The measurement matrix is defined as a random subset of rows of the full STOne transformation. Additionally, the measurement matrix defines the patterns to be projected in order to obtain the LTM measurements. The measurement patterns originally contain negative entries that cannot be projected using a digital projector. In Section IV, we provide more details with respect to overcoming the aforementioned limitation using complementary measurement patterns.

Undersampled high-resolution STOne coefficients can be reorganized into a complete set of low-resolution STOne transform coefficients. Let us assume that the high-resolution image has dimensions  $K \times K$  for simplicity (Fig. 2.1).



**FIGURE 2.** Visualization of STOne embedding. 1) The high-resolution image of size  $K \times K$ , where  $K = 8$ , is divided into  $\delta \times \delta$  blocks with  $\delta = 2$ . 2) Blocks of the high-resolution image can be rebinned into a low-resolution preview of size  $\kappa \times \kappa$ , where  $\kappa = 4$ . 3) Columnwise matrix vectorization. The dashed line denotes the border between two columns. 4) Visualization of the blockwise matrix vectorization, which facilitates the up-sampling process in the STOne LTM estimation. Dashed lines denote borders between the blocks.

The minimal number of measurements needed for the direct STOne preview is defined by the desired resolution of the preview image. To obtain a reconstruction at  $\kappa \times \kappa$  resolution (Fig. 2.2), one needs to acquire  $\kappa^2$  measurements. To be able to easily switch between the low-resolution preview and the high-resolution reconstruction, the authors in [9] propose the use of vector embedding, which is similar to nested-dissection embedding. The basic idea is to divide the original image into distinct blocks of size  $\delta \times \delta$  and to vectorize it block-by-block, instead of with the commonly used column-major ordering (Fig. 2.3). Blockwise vectorization (Fig. 2.4) guarantees that individual blocks are mapped contiguously into the vector. Additionally, such vectorization enables efficient up-sampling of the low-resolution previews using simple matrix operations (zero-order interpolation matrix).

In our work, STOne transformation is used for the design of the projection patterns for LTM measurement. Our main goal is to exploit direct STOne low-resolution previews for initial estimation of the nonzero coefficients in the LTM. An initial estimation of nonzero coefficients is used to reduce the search space of the sparse optimization algorithm. This approach significantly reduces the computational complexity and accelerates the LTM estimation while at the same time ensuring high accuracy of the LTM reconstruction. To our knowledge, this is the first application of STOne transformation to LTM estimation.

In our experiments, measurements  $\mathbf{c}_i$  are obtained by projecting patterns  $\phi_i$  defined by the rows of measurement matrix  $\Phi$  onto the observed scene and by capturing them using a camera. A set of  $k$  measurement patterns is sequentially projected onto the scene. After the measurement process is finished, a low-resolution LTM preview is calculated.

Low-resolution LTM has to be congruent with the obtained measurements and must satisfy:

$$\mathbf{c}_i = \Phi \mathbf{R} \mathbf{t}_i^L, \quad i = 1, \dots, (mn) \quad (9)$$

where  $\Phi$  is the measurement matrix whose rows contain the vectorized measurement patterns,  $\mathbf{R}$  is the zero-order interpolation operator that is used to upsample the low-resolution preview to the original size,  $\mathbf{t}_i^L$  is the low-resolution preview of the reflectance function for the  $i^{\text{th}}$  camera pixel, and  $\mathbf{c}_i$  is the corresponding measurement vector.

Equation 9 can be solved without using sparse optimization by applying the pseudoinverse to the system:

$$\mathbf{t}_i^L = ((\Phi \mathbf{R})^\top (\Phi \mathbf{R}))^{-1} (\Phi \mathbf{R})^\top \mathbf{c}_i. \quad (10)$$

The fact that the STOne transform is unitary and well conditioned enables simplification of the previous equation using:

$$\mathbf{t}_i^L = (\Phi \mathbf{R})^\top \mathbf{c}_i = \mathbf{R}^\top \Phi^\top \mathbf{c}_i, \quad (11)$$

where  $(\cdot)^\top$  denotes the Hermitian transpose operator. The system in (11) is under-determined, since  $k \ll (p \times q)$ . The solution obtained by the pseudoinverse minimizes the Euclidean norm, but does not yield a sparse solution. Usually, obtaining a sparse solution involves applying hard thresholding using a nonlinear operator  $\Gamma$ , which can be written in its binary form as:

$$\Gamma(x, \tau) = \begin{cases} 1 & \text{if } x > \tau \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

To obtain a sparse and binary low-resolution LTM preview  $\mathbf{t}_i^{LB}$ , one straightforward method would be to apply hard thresholding  $\Gamma$  to each entry  $t_{i,j}^L$  in  $\mathbf{t}_i^L$ :

$$t_{i,j}^{LB} = \Gamma(t_{i,j}^L, \tau), \quad j = 1, \dots, (pq). \quad (13)$$

LTM contains direct and global light transport components. Direct components of the LTM correspond to the light transport paths that have single reflections in the scene, while global components can include multiple bounces before reaching the camera sensor. Direct components have higher intensity than the global components and are well localized. Therefore, our thresholding algorithm is inspired by bilateral filtering, which takes into account both spatial and radiometric differences between the pixels in an image. First, we find the element with the largest amplitude in each  $\mathbf{t}_i^L$  (i.e., direct components), and denote it as  $t_{max}$ . After reshaping  $\mathbf{t}_i^L$  into a 2D image, spatial distance in the  $l_1$  norm can be defined as:

$$d_j = |x_j - x_{max}| + |y_j - y_{max}|, \quad j = 1, \dots, (pq), \quad (14)$$

where  $d_j$  denotes the distance of the  $j$ -th element in  $\mathbf{t}_i^L$  (with coordinates  $x_j, y_j$ ) to the highest amplitude element  $t_{max}$ , which is located at  $(x_{max}, y_{max})$ . We apply hard thresholding to the calculated distances using  $\Gamma(d_j, \tau_1)$ , where  $\tau_1$  denotes the distance threshold. Essentially, this approach sets all of the elements in  $\mathbf{t}_i^L$  where  $d_j > \tau_1$  to zero.

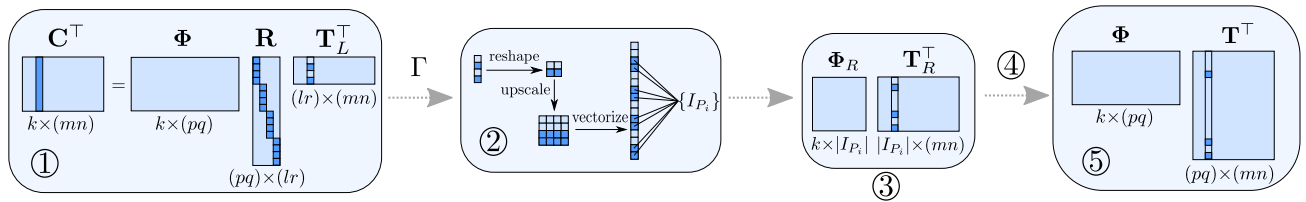


FIGURE 3. Flowchart of fast LTM estimation using STOne transformation. Algorithm 1 offers a detailed explanation of the flowchart.

Additionally, we search  $\mathbf{t}_i^L$  for elements that have amplitudes close to the amplitude of the highest intensity element. We can write this as  $\Gamma(A_j, \tau_2)$ , where  $A_j$  is the amplitude of the  $j$ -th element of  $\mathbf{t}_i^L$ , and  $\tau_2$  is the amplitude threshold related to the maximum intensity  $A_{max}$ .

Our final estimation of sparse and binary low-resolution LTM takes into account both spatial and amplitude filters using:

$$t_{i,j}^{LB} = \Gamma(d_j, \tau_1) \wedge \Gamma(A_j, \tau_2), \quad j = 1, \dots, (pq). \quad (15)$$

Using the described procedure, the low-resolution LTM enables obtaining a binary mapping that encodes the positions of potential nonzero coefficients in the high-resolution LTM. A shrunk measurement matrix  $\Phi_R$  and reduced LTM  $\mathbf{T}_R^T$  are obtained by applying the binary index map obtained in the previous step and removing the zero coefficients. After sparse optimization is applied in the reduced space, reconstructed LTM coefficients are rearranged into the full size LTM using the binary index map. A step-by-step description of the procedure is given in Alg. 1 and depicted in Fig. 3.

### C. DUAL IMAGING AND VIRTUAL RELIGHTING

The light transport matrix describes all necessary geometric and photometric relations within the camera-projector system for the observed scene. Equations 4 and 5 describe the primal and dual configurations of the system. In the primal configuration, light is emitted by the projector and captured by the camera according to their respective viewpoints. The transposed LTM describes the reversed light flow within the scene. It can be considered as if the roles of the camera and projector were interchanged.

LTM can be used to virtually capture images from the camera and projector viewpoints using:

$$\mathbf{c}' = \mathbf{T}\mathbf{1}' \quad \text{and} \quad \mathbf{p}'' = \mathbf{T}^T\mathbf{1}'', \quad (16)$$

where  $\mathbf{1}'$  and  $\mathbf{1}''$  correspond to vectorized white projector and camera images at their corresponding resolutions. Essentially, we virtually project uniform white illumination onto the scene in order to obtain the images in primal and dual configurations. The image captured by the camera in the primal configuration is shown in Fig. 4.1, and the image virtually captured by the projector in the dual configuration is shown in Fig. 4.2. It should be noted that both captures are obtained virtually, derived from (16), using the estimated light transport matrix.

### Algorithm 1 Fast LTM Estimation Using STOne Transformation

**INPUT:** Set of measurement vectors  $\mathbf{C}^T$ , measurement matrix  $\Phi$ , and size of the low-resolution STOne preview

**OUTPUT:** High-resolution light transport matrix  $\mathbf{T}^T$

**for**  $i = 1$  to  $mn$  **do**

① Estimate low-resolution light transport  $\mathbf{t}_i^L$  corresponding to the measurement vector  $\mathbf{c}_i$  using (11).

② Perform binarization of the low-resolution light transport  $\mathbf{t}_i^L$  using (15). Upscale  $\mathbf{t}_i^L$  to the original resolution of  $\mathbf{t}_i$  using interpolation operator  $\mathbf{R}$ . Create a subset of nonzero indices  $\{I_{P_i}\}$  from the upscaled binary light transport.

③ Reduce the search space for sparse optimization by removing zero coefficients in  $\mathbf{t}_i$  and obtain reduced light transport  $\mathbf{t}_i^R$ . Remove columns from the measurement matrix  $\Phi$  that include values of zero at the corresponding indices of the index map  $\{I_{P_i}\}$  and obtain the reduced measurement matrix  $\Phi_R$ .

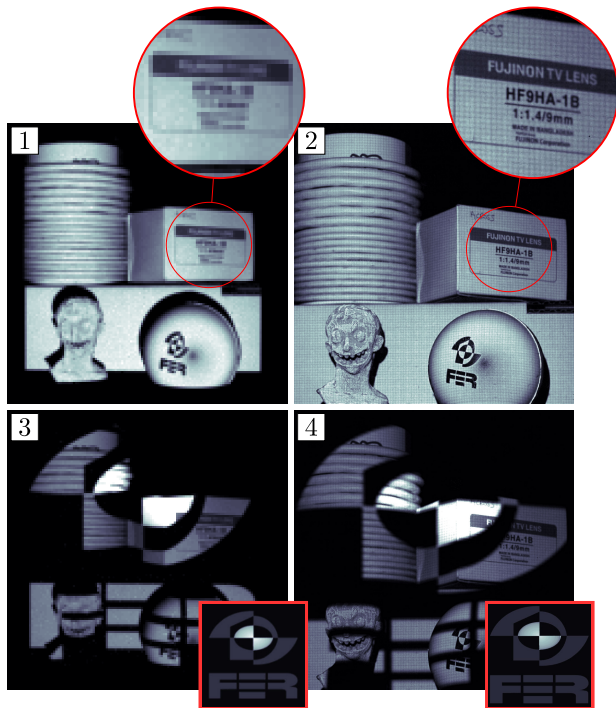
④ Perform sparse optimization as in (6) in the reduced search space and obtain the estimation of the reduced light transport matrix  $\mathbf{t}_i^R$ .

⑤ Reorder the coefficients of the reduced light transport  $\mathbf{t}_i^R$  into the high-resolution LTM using the previous mapping and obtain the high-resolution light transport estimate  $\mathbf{t}_i$ . Concatenate  $\mathbf{t}_i$  to the final high-resolution light transport matrix estimate  $\mathbf{T}^T$ .

**end for**

Previously, the scene was reilluminated using uniform illumination. Arbitrary illumination patterns can be projected using the same principles. In Fig. 4.3 and 4.4, virtual relighting in the primal and dual configurations using an arbitrary pattern is shown. We would like to emphasize that the aforementioned is performed as a postprocessing step and does not require additional measurements using the physical setup.

One of the key observations is that the spatial resolutions of the camera and projector are preserved in the dual configuration. In [7], the authors use a camera-projector system with matching resolutions. The final resolution of their 3D reconstruction is limited by the resolution of the imaging sensor. In contrast, in our research, we use a low-resolution camera in combination with a higher resolution

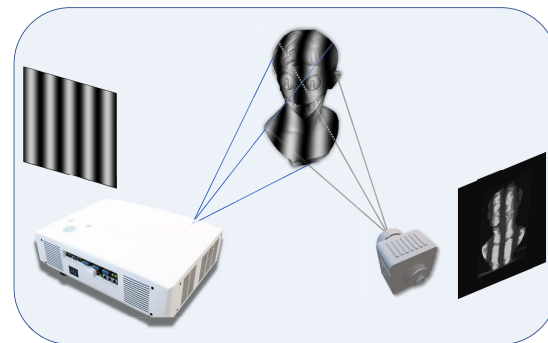


**FIGURE 4.** (1) Camera image in the primal configuration under uniform white projector illumination and (2) virtual projector “image” in the dual configuration under virtual white camera illumination. Notice the difference in the resolution and the viewpoint between the images: (3) virtual relighting of the scene in primal configuration and (4) virtual relighting in dual configuration. Projected patterns are shown in red rectangles at the bottom right corners of the subfigures. None of the images shown were captured using the physical measurement setup and instead were captured using the LTM and dual imaging principles.

projector. Low-resolution sensors are commonly used in imaging setups in nonvisible regions of the electromagnetic (EM) spectrum, and the cost of production for even such low-resolution sensors is significant. Our setup enables acquiring images of the observed scene from the projector viewpoint at the resolution of the digital micromirror device (DMD). In the following sections, we adapt structured light techniques in order to obtain high-resolution 3D reconstructions in the dual configuration of the camera-projector system.

### III. 3D RECONSTRUCTION USING DUAL IMAGING

A system comprising a paired camera and projector is usually used as a 3D scanning solution based on active vision, i.e., the structured light (SL) principle. In structured light, a 3D surface reconstruction is obtained by finding correct correspondences between the pixels of the coded projected pattern and the pixels captured with the camera (Fig. 5). One of the most popular SL methods is fringe projection profilometry (FPP) due to its robustness and high accuracy [10]. Furthermore, if the object of interest is static, the method of choice is the temporal phase shifting strategy described below.



**FIGURE 5.** Structured light approach for 3D reconstruction. The projector projects a coded pattern, and the camera captures the pattern distorted by the 3D surface. Finding the correct pixel correspondences between the projected pattern and the pattern captured by the camera enables triangulation and 3D reconstruction of the observed object.

#### A. PHASE SHIFTING PROFILOMETRY

In temporal phase shifting, a projector projects a set of  $N$  phase-shifted sinusoidal fringes:

$$I_n^p(x^p, y^p) = \frac{1}{2}I_0(1 + \cos(2\pi f^p x^p + 2\pi n/N)), \quad n = 0, \dots, N - 1, \quad (17)$$

where  $n$  denotes frame numbers (time-shifts),  $f^p$  the fringe frequency,  $I_0$  the projector’s intensity, and  $(x^p, y^p)$  the pixel coordinates of the projector.

The camera captures modulated projected intensities reflected by the object (directly and indirectly), together with the additional ambient light. The detailed notation of this intensity modulation is given in [11]; in this paper, as in most of the FPP literature, we use a simplified notation expressed as:

$$I_n^c(x^c, y^c) = I_A + \frac{1}{2}hI_0(1 + \cos(\Phi + 2\pi n/N)). \quad (18)$$

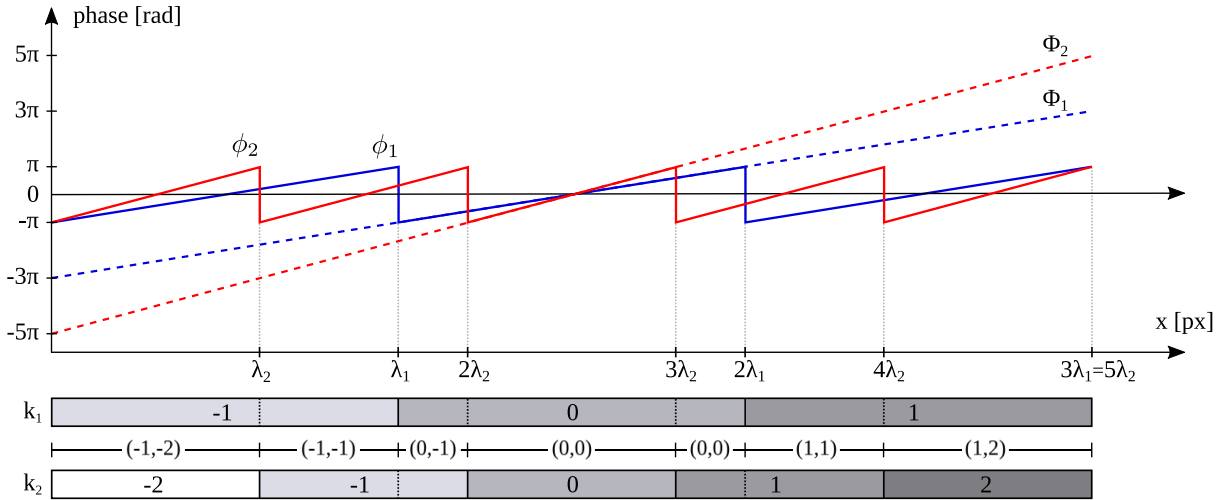
Variable  $I_A$  represents ambient illumination,  $h$  models intensity loss, and  $\Phi = 2\pi f^p x^p$  is the *spatial phase* that encodes the projector coordinates required for 3D reconstruction:

$$x^p = \frac{\Phi}{2\pi f^p}. \quad (19)$$

Here, without loss of generality, we assumed that the fringes are vertical (perpendicular to the  $x^p$ -axis), which means that only the  $x^p$  coordinate is *phase-encoded*. To encode the  $y^p$  coordinate, the same procedure is repeated with horizontal fringes.

After acquiring all  $N$  phase-shifted fringe images, one can estimate phase values measured modulo- $2\pi$ , i.e.,  $\phi \equiv \Phi(\text{mod } 2\pi)$  using:

$$\phi(x^c, y^c) = \text{atan2} \left( - \sum_{n=0}^{N-1} I_n^c(x^c, y^c) \sin(2\pi n/N), \sum_{n=0}^{N-1} I_n^c(x^c, y^c) \cos(2\pi n/N) \right). \quad (20)$$



**FIGURE 6.** Illustration of number-theoretical phase unwrapping approach in the case of two sinusoidal pattern sets. The first set is a low-frequency set, with  $f_1 = 3$ , and the second set is a high-frequency set, with  $f_2 = 5$ . Corresponding unwrapped phases are denoted with  $\Phi_1$  and  $\Phi_2$ . Coefficients  $k_1$  and  $k_2$  denote regions of  $\pm 2\pi$  jumps (addition or subtraction) required for the successful unwrapping of phases  $\phi_1$  and  $\phi_2$ . The main idea of this number-theoretical approach is to find the correct unique  $(k_1, k_2)$  pair from (pixelwise) measured  $\phi_1$  and  $\phi_2$  values using a predefined lookup table.

For a complete estimation of the spatial phase  $\Phi$ , the phase  $\phi$  must be unwrapped. This can be achieved using a number of different unwrapping approaches such as encoding the fringe order with color-coding or a statistical pattern, using additional pattern projection (binary coded pattern or statistical pattern), or by projecting an additional set of sinusoidal fringes with a different spatial frequency [10], [12]. The latter approach is called a multifrequency phase shift (MPS), and it is a robust and accurate approach when temporal unwrapping is possible, i.e., when the object of interest is static during pattern projection. Depending on the number of frequencies used and their mutual relation, MPS unwrapping methods can be classified into three groups [12]: hierarchical methods, heterodyne methods, and number-theoretical methods.

In our research, we decided to use the number-theoretical approach with a lookup table [13], [14]. Using this approach, one can successfully compute the spatial phase  $\Phi$  up to the value equal to  $\text{LCM}(\lambda_1, \lambda_2)$ , where  $\text{LCM}(\cdot, \cdot)$  represents the least common multiple of the wavelengths  $\lambda_1$  and  $\lambda_2$  of the corresponding projected sinusoidal patterns. If the wavelengths are chosen as prime numbers, LCM becomes a simple multiplication, and consequently  $f_1 = \text{LCM}(\lambda_1, \lambda_2)/\lambda_1 = \lambda_2$  and  $f_2 = \text{LCM}(\lambda_1, \lambda_2)/\lambda_2 = \lambda_1$ .

Figure 6 shows the basic principle of the chosen MPS phase unwrapping method. Let the chosen frequencies of the sinusoidal patterns be  $f_1 = 3$  and  $f_2 = 5$ . The unwrapped spatial phases are then equal:

$$\begin{cases} \Phi_1(x, y) = \phi_1(x, y) + 2\pi k_1(x, y) \\ \Phi_2(x, y) = \phi_2(x, y) + 2\pi k_2(x, y). \end{cases} \quad (21)$$

It should be evident that relation  $\Phi_1/\Phi_2 = f_1/f_2$  between two phases holds. From this relation and (21), it follows that:

$$k_2 f_1 - k_1 f_2 = \frac{f_2 \phi_1 - f_1 \phi_2}{2\pi}. \quad (22)$$

In most cases, the wavelengths are chosen as integers (in pixels), which means that the left side of (22) is an integer. Consequently, the right side of the equation also must be the same integer. This means that one can create a lookup table with predefined  $(k_1, k_2)$  pairs and corresponding values  $k_2 f_1 - k_1 f_2$ .

Once the wrapped phase values  $\phi_1(x, y)$  and  $\phi_2(x, y)$  are estimated using (20), the lookup formula:

$$(k_1, k_2) = \text{LUT} \left\{ \text{round} \left[ \frac{f_2 \phi_1 - f_1 \phi_2}{2\pi} \right] \right\} \quad (23)$$

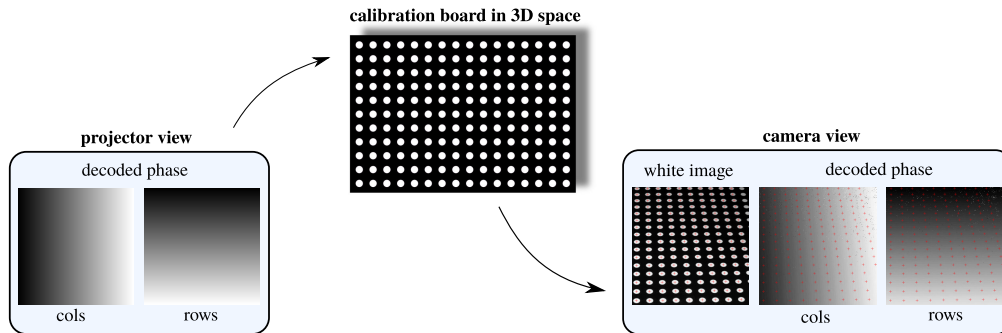
is used to determine the unique pair  $(k_1, k_2)$ . Finally, the unwrapped spatial phase  $\Phi$  can be computed using (21) and scaled with corresponding  $\lambda$  values to a valid range  $[0, 2\pi f]$ .

### B. CAMERA-PROJECTOR GEOMETRIC CALIBRATION

To reconstruct 3D positions of the captured pixels, it is necessary to first perform a geometric calibration of the camera-projector system. Since a projector is basically an inverse camera, it is possible to use a regular calibration procedure for a stereo (camera) pair with a few modifications.

In stereo systems with two cameras, the most common procedure for geometric calibration includes capturing multiple images of a flat checkerboard pattern positioned at different locations in the 3D space. When the distances between corner points of the checkerboard are known, by detecting those corner points in camera images and finding their mutual correspondences, it is possible to compute the intrinsic and extrinsic camera parameters via bundle adjustment [15]. Those parameters can be used later to triangulate the 3D locations of corresponding camera pixels in 3D space.

When one of the cameras in the stereo pair is replaced with a projector, one cannot directly match corresponding



**FIGURE 7.** Visualization of the calibration procedure for the camera-projector system. A calibration board with white circles is used and placed in the common FOV of the camera and projector. White image and MPS patterns are projected by the projector and captured using the camera. Pixel coordinates of the circle centers are estimated from the white image, and phase is decoded column- and rowwise from the MPS images. On the right, we demonstrate how camera pixels  $(c_x, c_y)$  are detected on the white image of the calibration board, and projector pixels  $(p_x, p_y)$  are extracted from the decoded phase over projector columns  $(p_x)$  and rows  $(p_y)$  at the corresponding circle center positions.

pixels of the calibration board because it is impossible to observe the calibration board directly from the projector view. Hence, we employ the MPS approach to find correspondences between pixels, an approach similar to the one presented by Moreno and Taubin [16].

We use a calibration board with white circles on a black background (Fig. 7). Correspondences between the projector pixels and the camera pixels are found via circle centers of the calibration board in the following manner:

- 1) A calibration board is placed somewhere in the common field-of-view (FOV) in front of the camera-projector system.
- 2) A white image and MPS patterns are projected using the projector, and corresponding images are captured with the camera.
- 3a) Pixel coordinates of circle centers  $(c_x, c_y)$  in the binarized white image of the calibration board are estimated using an appropriately chosen method. In our experiments, we use the MATLAB function `regionprops` [17]. This function works on a binary input image. First, pixels belonging to a certain region are found and labeled. Then, the centroid of each region is calculated by taking the mean of the  $x$  and  $y$  coordinates in each region. Appropriate filtering is applied to suppress outliers that do not correspond to the centers of white circles.
- 3b) The unwrapped spatial phase  $\Phi$  is computed for both horizontal and vertical directions, and phase values are scaled to match projector pixel resolution (width and height, respectively).
- 4) Each camera pixel  $(c_x, c_y)$  is paired with the values  $(p_x, p_y)$ , where  $p_{(\cdot)}$  denotes the decoded phase values (decoded over columns and rows, respectively) at the pixel  $(c_x, c_y)$ , i.e., the circle center.

This process is repeated for  $N > 2$  positions of the calibration board in 3D space, which are chosen in a way that a calibration board covers most of a 3D volume

of interest. The rest of the calibration procedure is the same as for the regular stereo camera pair using pixel correspondences  $(c_x, c_y) \longleftrightarrow (p_x, p_y)$ . Intrinsic and extrinsic parameters of the whole system are calculated using the estimated correspondences and built-in MATLAB function `estimateCameraParameters`. This function implements the standard Zhang's camera calibration method [18]. Zhang's method computes homographies between the points on the image plane and the calibration plane for each calibration position. Initial estimates of intrinsic and extrinsic parameters are then refined using the Levenberg-Marquadt nonlinear optimization algorithm.

### C. 3D RECONSTRUCTION IN DUAL WORLD

As explained in Section II-C, the LTM enables dual imaging, i.e., switching roles of the projector and the camera. Moreover, as mentioned in Section III-A, the MPS approach enables one to successfully reconstruct the 3D position of each camera pixel, provided that it belongs to the part of the scene illuminated by the projector.

To obtain 3D reconstruction in the dual configuration, we need to virtually project the sequence of  $N$  phase-shifted patterns (as defined in (17)) from the camera viewpoint onto the observed scene and “capture” the images of the modulated scene from the projector viewpoint. This can be written as:

$$\mathbf{p}_n'' = \mathbf{T}^\top \mathbf{c}_n'', \quad n = 0, \dots, N - 1, \quad (24)$$

where  $\mathbf{p}_n''$  denotes the images of the scene (modulated by the MPS projection) from the projector viewpoint, and  $\mathbf{c}_n''$  denotes the MPS patterns virtually projected by the camera. We would like to emphasize that the resolution of the projection device is circumvented in fringe profilometry using MPS since the phase values are estimated and matched at subpixel level. In the case when the projector has higher resolution than the camera, dual imaging is a suitable method for obtaining



3D reconstruction at a higher resolution than the physical resolution limit of the imaging device.

The previously described process results in 3D reconstruction in the dual configuration of the camera-projector system. Similarly, to obtain 3D reconstruction in the primal configuration, the virtual MPS measurement process must be performed using:

$$\mathbf{c}'_n = \mathbf{T}\mathbf{p}'_n, \quad n = 0, \dots, N - 1. \quad (25)$$

#### IV. EXPERIMENTAL RESULTS

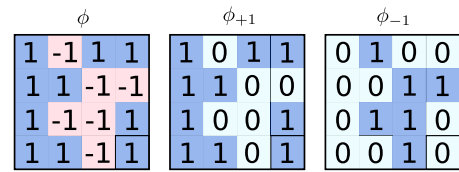
##### A. MEASUREMENT SETUP

Our measurement setup is composed of an off-the-shelf InFocus IN3118HD DLP projector and Smartek GC2441C camera. We closely follow the calibration procedure for camera-projector systems proposed in [19]. For LTM reconstruction and experiments in dual imaging, we use an Intel Core i7-4930K CPU@3.40 GHz with 64 GB of RAM.

The camera-projector system has to satisfy several prerequisites. One-on-one mapping has to be achieved between the computer generated measurement patterns and the projected patterns. The projector must operate in native resolution, and the resolution of the PC video controller has to match the native resolution of the projector in order to avoid potential spatial resampling. Projector manufacturers commonly reduce image bandwidth by using chroma subsampling. In chroma subsampling, the resolution of color channels is decimated, while the resolution of the luminance channel is preserved. Undesired intensity variations occur when projecting patterns with higher resolution than the resolution of the color channels. IN3118HD uses 4:2:2 chroma subsampling; thus, we project patterns that are  $2\times$  downsampled to match the chroma subsampling ratio. Although this approach reduces the effective projector resolution, it offers the benefits of increased signal-to-noise ratio of the overall system.

Furthermore, the camera has to operate in linear mode, which is ensured by obtaining raw images directly from the sensor. Camera exposure time and projector color-wheel rotation have to be synchronized. We adjust the camera exposure time to be a multiple of the projector color-wheel rotation period in order to avoid potential aliasing caused by the color-wheel rotation. The color-wheel of the projector operates at 120 Hz, and we thus set the camera exposure time to  $1/60$  s [19]. While synchronization within our measurement system is based on timers and basic triggers, there are more advanced software [20] or hardware solutions [21] that could further improve the acquisition speed of the camera-projector system.

To estimate the light transport matrix, we project STOne measurement patterns defined by the measurement matrix  $\Phi$  onto the scene. The IN3118HD projector is an 8-bit projector that can project only values within the  $[0, 255]$  range. Rows of the STOne transformation matrix define the measurement patterns  $\phi$  and contain  $\pm 1$  entries. Since the projector can project only in the  $[0, 255]$  range, we need to split the original



**FIGURE 8.** Example of a measurement pattern  $\phi$  that contains  $\pm 1$  entries. Since the projector can project only positive components, the original  $\phi$  is split into  $\phi_{+1}$  and  $\phi_{-1}$  that are complementary to each other and contain only 1 and 0 entries, which are respectively mapped to 255 and 0 within the dynamic range of the projector.

measurement patterns  $\phi$  into positive  $\phi_{+1}$  and negative  $\phi_{-1}$  components (see Fig. 8). In  $\phi_{+1}$ , positive entries are mapped to value 255, while entries of zero are mapped to value 0 in the projector dynamic range. In  $\phi_{-1}$ , the entries are complements of  $\phi_{+1}$ .

Single measurement  $\mathbf{c}_i$  is then defined as the difference between the images of the scene when illuminated by positive and negative patterns:  $\mathbf{c}_i = \mathbf{c}_i^{+1} - \mathbf{c}_i^{-1}$ . This approach eliminates the influence of projector background light and reduces the noise in the measurements.

In addition to the previously described calibration, we perform geometric calibration of the camera-projector system. The planar calibration board has to be positioned at several positions, and images of projected patterns (i.e., white and MPS) have to be captured. Note that the calibration board positions should cover the entire working volume. In our calibration, we acquire calibration images for a total of 20 board positions. Geometric calibration is performed according to the procedure described in Section III-B. Estimated calibration parameters are used for 3D reconstruction in both primal and dual worlds.

Accuracy of geometric calibration is crucial for the overall performance of the 3D reconstruction using camera-projector systems. Although the calibration procedure can be time-consuming, it is an offline procedure that has to be performed only once under the assumption of fixed geometry of the camera-projector system.

##### B. EXPERIMENTS

Our first goal is to compare the proposed method to the competitive LTM estimation methods in terms of time complexity and reconstruction accuracy. The proposed method can be considered as a special case of the conventional CS LTM estimation. The method proposed by Chiba and Hashimoto in [7] is most similar to the proposed method. Therefore, we perform a comparison between the proposed solution, Chiba’s approach and the full solver LTM estimation approach. The resolutions of the camera and projector in this setup are  $(128\times 128)$  and  $(256\times 256)$ , respectively. LTM reconstructions are performed for two different scenes, and time complexity and reconstruction quality are considered.

To create a fair comparison framework, we use the SPAMS optimization toolbox [22] for solving the sparse optimization problem and use the unconstrained form as in (6). Parameter  $\lambda$  is set to  $1e-3$ . The total number of measurements is set

to  $M = 1024$  for all of the methods. In the implementation of Chiba's method, the total number of measurements is split between 4 different resolutions in the projected resolution pyramid. We closely follow the approach described in [7] and increase the total number of the measurement patterns to  $M = 1024$ . Thus, we project the random Gaussian measurement patterns in 4 resolutions, ( $8 \times 8$ ,  $32 \times 32$ ,  $128 \times 128$ , and  $256 \times 256$ ), with the number of patterns per resolution being (32, 256, 480, 256), respectively. One additional measurement is taken for the projector background light estimation. For more details on the implementation of Chiba's method, we advise the interested reader to consult [7].

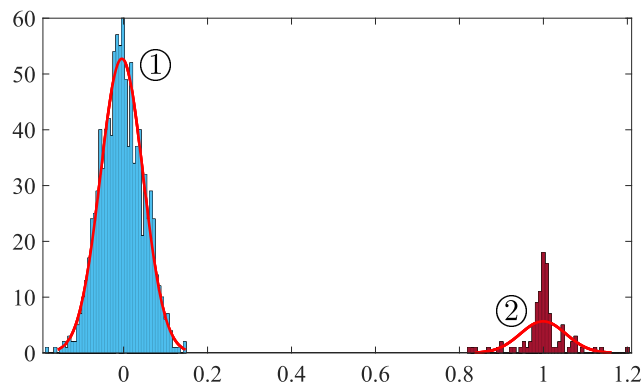
In the second part of our experiments, the main goal is to demonstrate the efficiency of 3D reconstruction in the dual configuration of the camera-projector system using the proposed method. A scene consisting of several objects (i.e., small figurine, ball, pencil holder with cables wrapped around it and a small box with inscription), as shown in Fig. 4, is used for this experiment. We analyze different aspects of the proposed system, including the accuracy and time complexity of the LTM estimation and 3D reconstruction.

After the measurement acquisition, LTM reconstruction closely follows the procedure described in Section II-B. In this experiment, the projector resolution is set to  $512 \times 512$ , while the camera resolution is reduced to  $128 \times 128$  pixels. STOne previews have  $\delta = 16$  times lower resolution than the original LTM in both horizontal and vertical directions. For the low-resolution STOne preview to be feasible, the Nyquist measurement rate has to be satisfied. We reconstruct  $32 \times 32$  STOne previews; thus, a minimum of 1024 measurement patterns is needed. In our experiments, we use 1% of the total number of measurements (i.e., 2622) required by the "brute-force" scanning algorithm.

Locations of the nonzero elements in the low-resolution LTM are estimated using (15). In the following, we provide general guidelines for the selection of the spatial and amplitude thresholds. The bidirectional reflectance distribution function (BRDF) is a function that precisely defines how light is reflected at a surface of the observed 3D object. Instead of modeling the complete BRDF, we use a reasonable approximation wherein the reflected beam is scattered in a  $\pm 10^\circ$  cone, which corresponds to approximately  $\tau_1 = 7 * \delta = 112$  pixel distance on the sensor. Spatial thresholding results in an increased sparsity of the LTM and reduces the computational cost.

In addition to the spatial threshold, we use an amplitude threshold. In Fig. 9, a visualization of a characteristic amplitude histogram for a single projector reflectance function is shown. Notice the bimodal character of the distribution: the noise component is on the left-hand side, and the useful information is on the right-hand side. From the visualization, it is clear that a wide range of amplitude thresholds would result in correct classification of the elements.

Sparse optimization is then performed in the reduced search space in order to obtain the high-resolution LTM. All computation is performed in sparse arithmetic to reduce



**FIGURE 9.** Visualization of the bimodal distribution of the LTM coefficients in the STOne preview of the projector reflectance function. Projector pixels corresponding to the illuminated camera pixel have significantly higher mean values than the noise level, resulting in a bimodal histogram. It is clear that a wide range of thresholds results in correct classification of the elements.

**TABLE 1.** Time complexity of the LTM reconstruction using different methods (in seconds).

TIME (s)	scene #1	scene #2
full solver	4.7163e+04	2.9982e+04
Chiba	165.91	151.53
STOne (proposed)	289.03	265.27

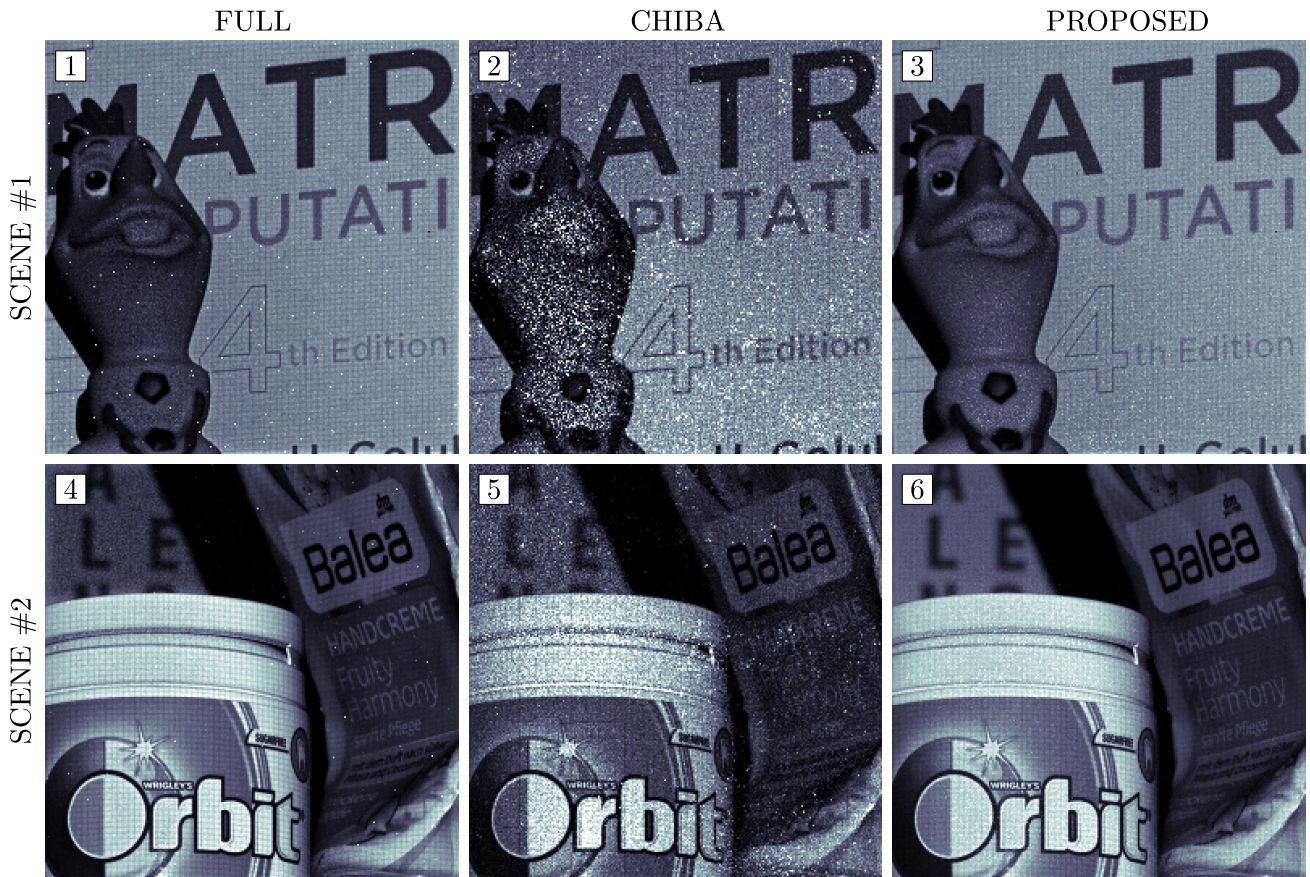
the memory footprint. In the final step of rapid LTM estimation using the proposed procedure, we again use the SPAMS optimization toolbox to solve the sparse optimization problem defined in (6).

### C. RESULTS

In this section, we provide a comparison of the results for the full LTM estimation, Chiba's method and the proposed method. First, we compare the time complexities of the algorithms. Acquisition of the  $M = 1024$  patterns using our measurement system requires approximately ten minutes, and the reported time is the same for all of the compared methods. It should be noted that synchronization of the camera and projector within our setup is performed using basic triggers and timers. Better synchronization methods could significantly reduce the acquisition times.

In Table 1, we report the LTM reconstruction times for the compared algorithms and for two different scenes. LTM reconstruction using sparse optimization in the full solution space requires several hours to finish. Chiba's method and the proposed method are comparable in terms of the reconstruction times, although Chiba's method is slightly faster. Reconstruction time using the full solver depends on the number of nonzero coefficients in the LTM, and therefore, the difference between the reconstruction times for scenes #1 and #2.

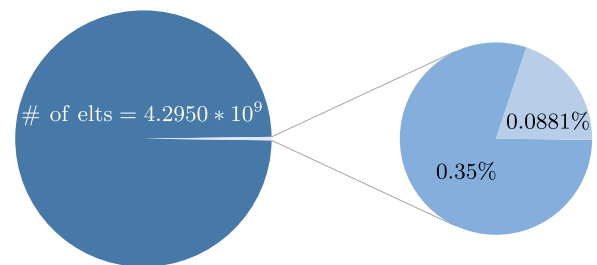
Next, we conduct a comparison of the three methods in terms of the reconstruction quality. In Fig. 10, reconstructions of the dual images are shown. Notice that the proposed



**FIGURE 10.** Comparison of two scene reconstructions in the dual configuration of the camera-projector setup. Columns correspond to the scene reconstructions obtained using full solver, Chiba’s method, and the proposed method. The proposed LTM estimation method yields similar or slightly better results than the full solver method in terms of the reconstruction quality, requiring only a fraction of the time. Moreover, the proposed method drastically outperforms Chiba’s method in terms of the reconstruction quality, with only a negligible increase in the computational cost.

method and the full LTM estimation methods yield similar results. Reconstructions obtained using the full solver contain slightly more noise (both impulse and Gaussian) than the proposed method. We argue that this additional noise is due to the unconstrained solution space in the reconstruction process in the original approach. Chiba’s method yields significantly noisier reconstructions than the proposed method does. Performed experiments confirm that the proposed method for solution space reduction in the LTM estimation is efficient.

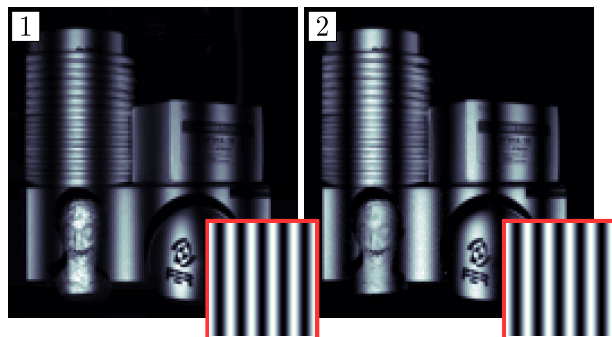
The experimental results for the second part of the experiments are reported in the sequel. We report the results that confirm the efficiency of the proposed search space reduction method. While most of the conventional CS LTM estimation algorithms work in high-dimensional space, the basic idea of the proposed STOne LTM estimation method is to reduce the search space of the sparse optimization algorithm. In Fig. 11, a pie chart visualization of nonzero elements in LTM estimation is shown. The total number of elements in LTM is defined by the camera and projector resolutions and equals  $4.2950 \cdot 10^9$  elements in our setup. The total number of nonzero elements obtained from the low-resolution LTM using the proposed algorithm is approximately 0.35% of the



**FIGURE 11.** Camera and projector resolutions define the size of LTM. We use  $512 \times 512$  projector resolution and  $128 \times 128$  camera resolution, which correspond to a total of  $4.2950 \cdot 10^9$  elements. The low-resolution STOne preview enables reduction of the search space to only 0.35% of the original space. Finally, only 0.0881% of the total number of elements are nonzero in the high-resolution LTM estimation.

total number of elements in the LTM. After applying sparse optimization to the initial estimate, only 0.0881% of the nonzero elements remain in the LTM.

To verify the accuracy of the LTM estimated using the proposed method, we perform several experiments. The basic idea is to compare the images obtained by the physical camera-projector system and the images obtained in the

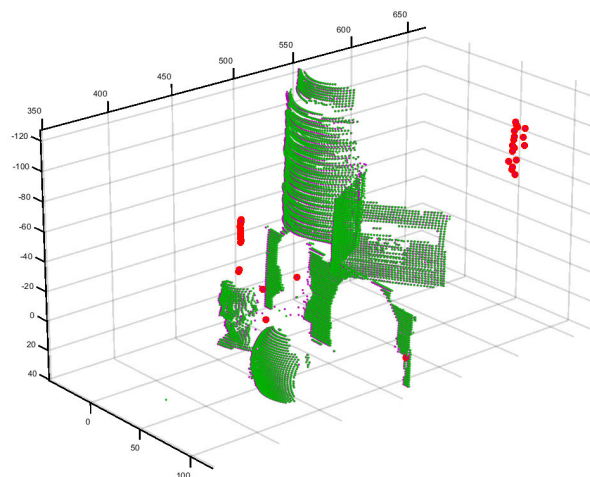


**FIGURE 12.** Comparison of real-world and virtual camera images illuminated by a fringe pattern. (1) represents a camera image captured using the real measurement setup, while (2) is an image obtained by virtually projecting a fringe pattern onto the scene and capturing it.

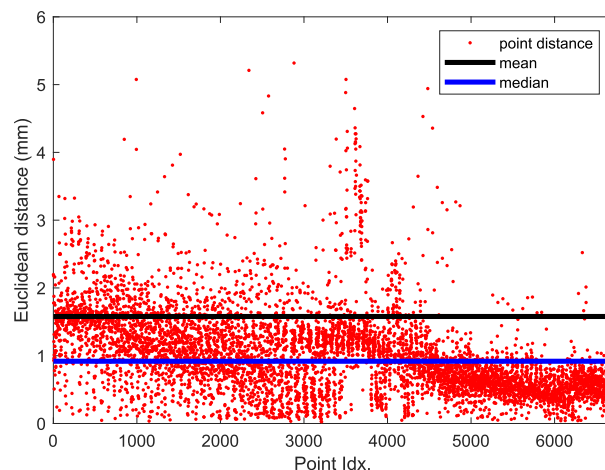
primal configuration using the estimated LTM. If our LTM reconstruction is accurate, images obtained by the real-world measurement setup and the images obtained using virtual projection in the primal configuration using (4) should be similar. In Fig. 12, a comparison of images obtained using the real-world and virtual camera-projector setup is shown. In both configurations, the projector illuminates the scene using a single MPS pattern while the camera captures the image of the modulated scene. Notice that the difference between the images is negligible, which confirms that our LTM accurately models the light transport within the observed scene.

To quantify the aforementioned results, we perform 3D reconstruction using the MPS measurements obtained using the physical camera-projector setup and compare it to the 3D reconstruction obtained from the MPS measurements obtained using virtual projection based on the estimated LTM. Figure 13 shows a comparison of the two point clouds. Reconstruction obtained using real measurements produces several outliers, while outliers do not occur in the virtual reconstruction. Reconstructed points that exist outside the observed volume are considered outliers and are visualized using the red points in Fig. 13. It is well known that the accuracy of fringe profilometry is susceptible to global illumination [6], [23], while the global illumination effects might be reduced in the estimated LTM. In our experiments, a point is considered an outlier if the distance between two corresponding points in the real and virtual 3D reconstruction is larger than 5 mm. In Fig. 14, we show the Euclidean distances between corresponding points in the 3D reconstructions that are obtained using the same parameters. The median distance between the two reconstructions is less than 1 mm, which confirms the efficiency of the proposed method.

Our final goal is to improve the resolution of the 3D reconstruction using virtual MPS measurement in the dual camera-projector configuration. Transposition of the LTM enables interchanging the roles of the camera and projector. Thus, we virtually project fringe patterns using the camera and “capture” them using the projector. We apply the previously used method for 3D reconstruction with the same parameters. Figure 15 shows a comparison of 3D reconstructions in primal and dual worlds of the camera-projector system.



**FIGURE 13.** Comparison of real-world and virtual point cloud reconstructions. Distances between corresponding points in the point clouds are small for most of the reconstructed pairs, although there are some outliers in the reconstruction obtained using the real measurement setup. Points that lie outside the observed volume are considered outliers and are denoted using red markers in the visualization.

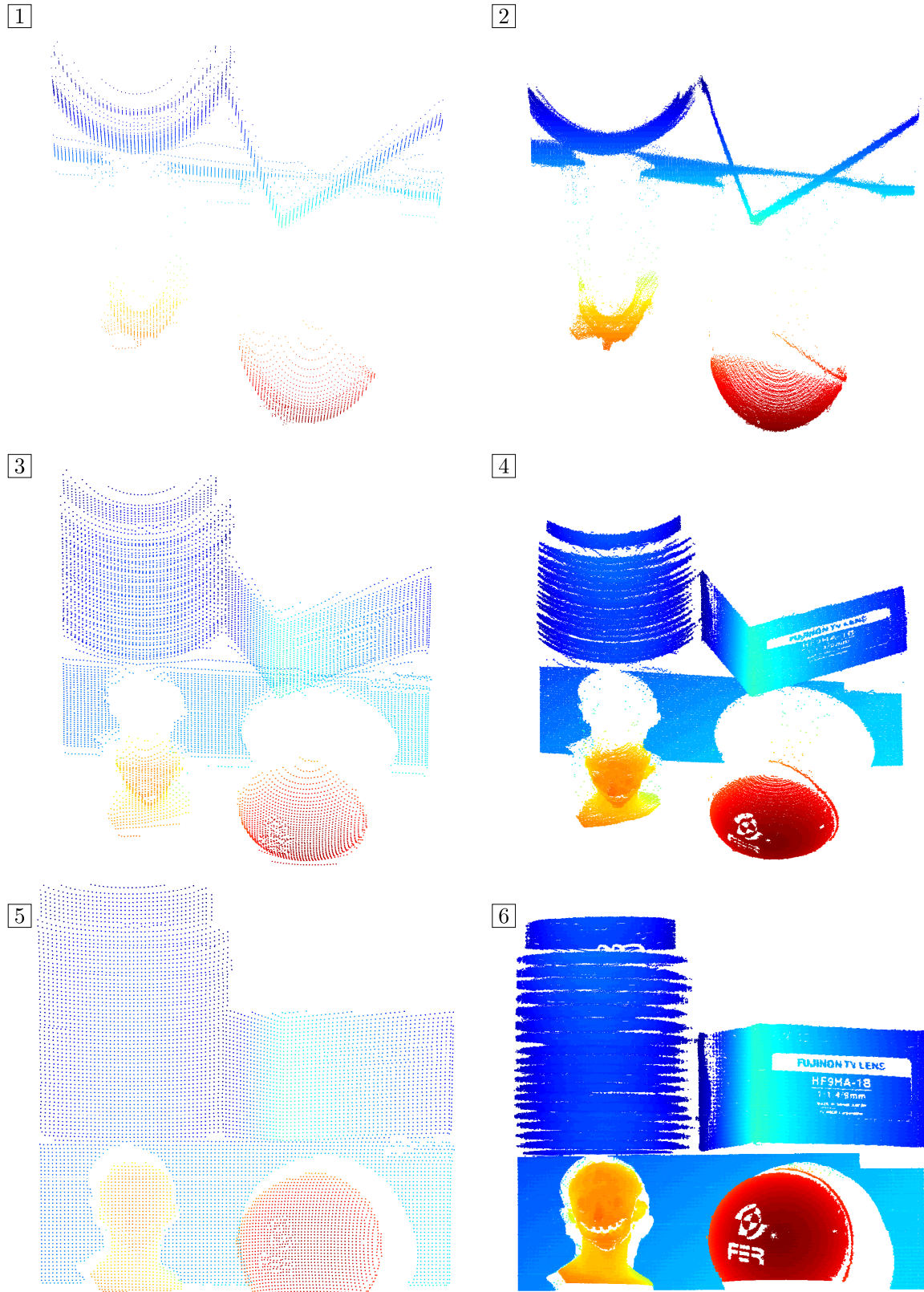


**FIGURE 14.** Euclidean distance plot between corresponding 3D points reconstructed from real MPS measurements obtained using the physical camera-projector setup and virtual MPS projection using the LTM in primal world. The mean and median distances are denoted by black and blue lines, respectively. Extreme outliers have been ignored in the visualization. The median distance between the two reconstructions is less than 1 mm.

Obviously, the point cloud obtained by the dual configuration is significantly denser and preserves much more detail (see the supplementary materials for better visualization and additional scene reconstructions).

## V. DISCUSSION

In our research, we successfully addressed two advances. First, we proposed an efficient method for LTM estimation based on STOne transformation, which significantly reduces the reconstruction time from the order of several hours to the order of minutes while achieving high accuracy of the light transport model. Second, we performed high-resolution 3D reconstruction by utilizing dual imaging and fringe profilometry techniques.



**FIGURE 15.** 3D reconstruction in primal (1,3,5) and dual (2,4,6) camera-projector configurations. Note that the point cloud obtained in the dual camera-projector configuration is significantly denser and contains much more detail.

Efficient light transport estimation reduces the existing gap between light transport modeling for research purposes and practical applications. Some of the potential applications of the proposed method are radiometric compensation of camera-projector systems and rapid dual imaging. Dual imaging is particularly interesting in imaging modalities that use imaging sensors that operate in nonvisible regions of the electromagnetic spectrum. Our method offers an ability to reduce the cost of production for high-resolution imaging sensors. A similar observation holds for 3D imaging in the dual camera-projector configuration. For example, in [24], the authors use a SWIR camera that has  $320 \times 256$  resolution for 3D face measurement. Our method can potentially improve the imaging resolution of the camera as well as the resolution of the 3D reconstruction.

Additionally, current benchmarks for fringe profilometry methods use real camera-projector systems to compare the efficiency of the algorithms. To obtain fair comparisons, imaging needs to be performed under the same conditions, which is hard to achieve in real measurement setups. Dual imaging offers an elegant solution for this problem. After LTM acquisition, comparison of fringe projection methods can be performed virtually under exactly the same conditions, providing a fair comparison framework. We will address the aforementioned topics in our future research.

## VI. CONCLUSION

In this paper, we present a method for fast light transport matrix estimation based on sum-to-one transformation and compressive sensing principles. Dual imaging is a photographic technique based on modifying light transport properties of a scene and enables powerful post-acquisition processing such as virtual viewpoint change and virtual relighting. The light transport matrix enables efficient modeling of light transport in a camera-projector system within an observed scene and contains information on both optical systems and the observed scene. Comparison of the proposed LTM estimation method with the baseline and competitive methods confirms efficiency in terms of the computational complexity and reconstruction quality.

Furthermore, we present a method for efficient 3D reconstruction in the dual configuration of the camera-projector system, where virtual measurements of the 3D scene are performed instead of physical measurements. We show that 3D scene reconstruction from the proposed virtual measurements corresponds with the actual physical acquisition. Moreover, this approach provides much more detail in the reconstruction. The computational complexity of the proposed method is reduced to such a level that practical implementations are feasible.

## REFERENCES

- [1] H. Von Helmholtz, *Handbuch Der Physiologischen Optik*. Leipzig, Germany: Leopold Voss, 1867. [Online]. Available: <https://www.e-rara.ch/zut/doi/10.3931/e-rara-21259>
- [2] G. G. Stokes, "On the perfect blackness of the central spot in newton's rings, and on the verification of fresnel's formula for the intensities of reflected and refracted rays," in *Mathematical and Physical Papers*, vol. 2. Cambridge, U.K.: Cambridge Univ. Press, pp. 89–103, 2009, doi: [10.1017/cbo9780511702259.008](https://doi.org/10.1017/cbo9780511702259.008).
- [3] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch, "Dual photography," *ACM Trans. Graph.*, vol. 24, no. 3, p. 745, Jul. 2005, doi: [10.1145/1073204.1073257](https://doi.org/10.1145/1073204.1073257).
- [4] P. Sen and S. Darabi, "Compressive dual photography," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 609–618, Apr. 2009, doi: [10.1111/j.1467-8659.2009.01401.x](https://doi.org/10.1111/j.1467-8659.2009.01401.x).
- [5] P. Peers, D. K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi, and P. Debevec, "Compressive light transport sensing," *ACM Trans. Graph.*, vol. 28, no. 1, pp. 1–18, Jan. 2009, doi: [10.1145/1477926.1477929](https://doi.org/10.1145/1477926.1477929).
- [6] N. Chiba and K. Hashimoto, "3D measurement by estimating homogeneous light transport (HLT) matrix," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Aug. 2017, pp. 1763–1768, doi: [10.1109/icma.2017.8016084](https://doi.org/10.1109/icma.2017.8016084).
- [7] N. Chiba and K. Hashimoto, "Ultra-fast multi-scale shape estimation of light transport matrix for complex light reflection objects," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–5, doi: [10.1109/icra.2018.8460892](https://doi.org/10.1109/icra.2018.8460892).
- [8] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal  $\ell_1$ -norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006, doi: [10.1002/cpa.20132](https://doi.org/10.1002/cpa.20132).
- [9] T. Goldstein, L. Xu, K. F. Kelly, and R. Baraniuk, "The STOne transform: Multi-resolution image enhancement and compressive video," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5581–5593, Dec. 2015, doi: [10.1109/tip.2015.2474697](https://doi.org/10.1109/tip.2015.2474697).
- [10] S. Zhang, *High-Speed 3D Imaging with Digital Fringe Projection Techniques*. Boca Raton, FL, USA: CRC Press, Apr. 2016. [Online]. Available: <https://doi.org/10.1201/b19565>
- [11] C. Zuo, S. Feng, L. Huang, T. Tao, W. Yin, and Q. Chen, "Phase shifting algorithms for fringe projection profilometry: A review," *Opt. Lasers Eng.*, vol. 109, pp. 23–59, Oct. 2018, doi: [10.1016/j.optlaseng.2018.04.019](https://doi.org/10.1016/j.optlaseng.2018.04.019).
- [12] C. Zuo, L. Huang, M. Zhang, Q. Chen, and A. Asundi, "Temporal phase unwrapping algorithms for fringe projection profilometry: A comparative review," *Opt. Lasers Eng.*, vol. 85, pp. 84–103, Oct. 2016, doi: [10.1016/j.optlaseng.2016.04.022](https://doi.org/10.1016/j.optlaseng.2016.04.022).
- [13] C. E. Towers, D. P. Towers, and J. D. C. Jones, "Time efficient chinese remainder theorem algorithm for full-field fringe phase analysis in multi-wavelength interferometry," *Opt. Express*, vol. 12, no. 6, p. 1136, 2004, doi: [10.1364/opex.12.001136](https://doi.org/10.1364/opex.12.001136).
- [14] M. Wang, "Phase unwrapping by a lookup table method: Application to phase maps with singular points," *Opt. Eng.*, vol. 38, no. 12, p. 2075, Dec. 1999, doi: [10.1117/1.602314](https://doi.org/10.1117/1.602314).
- [15] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Vision Algorithms: Theory and Practice*. Berlin, Germany: Springer, 2000, pp. 298–372, doi: [10.1007/3-540-44480-7\\_21](https://doi.org/10.1007/3-540-44480-7_21).
- [16] D. Moreno and G. Taubin, "Simple, accurate, and robust projector-camera calibration," in *Proc. 2nd Int. Conf. 3D Imag., Model., Process., Vis. Transmiss.*, Oct. 2012, pp. 464–471, doi: [10.1109/3dimpvt.2012.77](https://doi.org/10.1109/3dimpvt.2012.77).
- [17] *MATLAB R2019a, Computer Vision Toolbox Version 9.0*, The MathWorks, Natick, MA, USA, 2019.
- [18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: [10.1109/34.888718](https://doi.org/10.1109/34.888718).
- [19] I. Ralasic, D. Sersic, and D. Petrinovic, "Off-the-Shelf measurement setup for compressive imaging," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 502–511, Feb. 2019, doi: [10.1109/tim.2018.2847018](https://doi.org/10.1109/tim.2018.2847018).
- [20] T. Petković, T. Pribanić, M. Đonlić, and N. D'Appuzo, "Software synchronization of projector and camera for structured light 3D body scanning," in *Proc. 7th Int. Conf. 3D Body Scanning Technol.*, Lugano, Switzerland, Nov./Dec. 2016, pp. 286–295, [Online]. Available: <https://doi.org/10.15221/16.286>
- [21] J. Takei, S. Kagami, and K. Hashimoto, "3,000-fps 3-D shape measurement using a high-speed camera-projector system," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct./Nov. 2007, pp. 3211–3216, doi: [10.1109/iros.2007.4399626](https://doi.org/10.1109/iros.2007.4399626).

- [22] J. Mairal, “Sparse modeling for image and vision processing,” *Found. Trends Comput. Graph. Vis.*, vol. 8, nos. 2–3, pp. 85–283, 2014, doi: [10.1561/06000000058](https://doi.org/10.1561/06000000058).
- [23] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, “Structured light 3D scanning in the presence of global illumination,” in *Proc. CVPR*, Jun. 2011, pp. 713–720, doi: [10.1109/cvpr.2011.5995321](https://doi.org/10.1109/cvpr.2011.5995321).
- [24] S. Heist, M. Landmann, M. Steglich, Y. Zhang, P. Kühmstedt, and G. Notni, “Pattern projection in the short-wave infrared (SWIR): Accurate, eye-safe 3D shape measurement,” *Proc. SPIE*, vol. 10991, May 2019, Art. no. 109910J, doi: [10.1117/12.2518205](https://doi.org/10.1117/12.2518205).



**IVAN RALASIC** (Student Member, IEEE) was born in Požega, Croatia, in 1992. He received the B.Sc. degree in computing and the M.Sc. degree in information and communication technology from the University of Zagreb, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Faculty of Electrical Engineering and Computing. His current research interests include signal processing, compressive sensing, sparse modeling, and machine learning.



**MATEA DONLIC** (Student Member, IEEE) was born in Zagreb, Croatia, in 1990. She received the B.Sc. degree in computing, the M.Sc. degree in information and communication technology, and the Ph.D. degree in computing from the University of Zagreb, in 2012, 2014, and 2019, respectively. Her current research interests include computer vision, focusing on image analysis, shape analysis, and 3D reconstruction.



**DAMIR SERSIC** (Member, IEEE) received the Diploma degree and the M.S. and Ph.D. degrees in electrical engineering from the University of Zagreb, Zagreb, Croatia, in 1986, 1993, and 1999, respectively. Since 1987, he has been with the Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently a Full Professor. His current research interests include theory and applications of wavelets, advanced signal and image processing, adaptive systems, blind source separation, and compressive sensing. He is also a member of the European Association for Signal Processing. From 2006 to 2008, he has served as the Chair for the Croatian IEEE Signal Processing Chapter.

...