Hybrid single-pixel camera for dynamic hyperspectral imaging

Thomas Maitre, Elie Bretin, L. Mahieu-Williame, Michaël Sdika, Nicolas Ducros

HYBRID SINGLE-PIXEL CAMERA FOR DYNAMIC HYPERSPECTRAL IMAGING

Thomas Maitre [1], Elie Bretin [2], Laurent Mahieu-Williame [1], Michaël Sdika [1], Nicolas Ducros [1, 3]

[1]Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne,
CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, LYON, France
[2]Univ Lyon, INSA de Lyon, CNRS UMR 5208, Institut Camille Jordan, F-69621 Villeurbanne, France
[3]Institut universitaire de France (IUF), France

ABSTRACT

Single-pixel imaging recovers an image from a sequence of point measurements that correspond to inner products between the image of the scene and some light patterns. This technique is particularly suited to hyperspectral imaging which has a wide range of biomedical applications, in particular fluorescence-guided neurosurgery where high spectral resolution is required to improve tumor resection. However, the sequential nature of the acquisition leads to strong artifacts when a dynamic scene is considered. This paper proposes a simple yet efficient strategy for imaging such dynamic scenes. Our method relies on a hybrid device that combines a standard imaging arm with a hyperspectral single-pixel imaging arm. By estimating the motion from the standard imaging arm, we show how to compensate the motion perceived by the hyperspectral arm. We demonstrate the efficiency of the proposed strategy on simulated and experimental data.

Index Terms— Image reconstruction, motion compensation, computational optics, single-pixel imaging.

1. INTRODUCTION

Single-pixel imaging (SPI) acquires an image by measuring a series of inner products between some light patterns and the scene [1], where the light patterns are typically obtained using a digital micro-mirror device (DMD) and only a single point detector is required. The concept of SPI easily generalizes to hyperspectral imaging by replacing the point detector with a spectrometer, thus acquiring a hypercube with thousands of spectral channels in approximately ten seconds [2]. Hyperspectral imaging is particularly relevant for biomedical applications as it provides a non-contact and non-ionizing sensing technique for disease diagnosis and image-guided surgery. Numerous studies have used hyperspectral imaging to detect cancers, cardiac disease, ischemic tissue, skin burn, retinal pathologies, diabetes, kidney disease, and more [3]. We are particularly interested in fluorescence-guided neurosurgery where gliomas can be distinguished from healthy tissue by the injection of a contrast agent (e.g. protoporphyrin-IX [4]), provided a high spectral resolution is available.

Single-pixel acquisitions lead to an inverse problem where the image of the scene needs to be reconstructed from the measurements. While most reconstruction methods assume the scene is static [5], fast-moving dynamic scenes lead to strong blurring artifacts. A widespread strategy to tackle this issue in medical imaging is motion compensation (e.g., [6], [7], [8]). In SPI, [9] estimates the motion from a low-resolution video to add an optical-flow constraint during the reconstruction. In [10], a dynamic reconstruction method is proposed assuming prior knowledge of a constant volume deformation. The ideas of [9] and [10] are combined in [11]. In [12], specific patterns are inserted in the sequence to estimate the motion. The work presented in [13] investigates the use of deep-learning for video reconstruction. Motion compensation strategies, however, often come with important restrictions: the videos reconstructed in [9] and [11] have a smaller frame rate than the number of measurements, while in [12], we consider a rigid object and a static background.

In this paper, we propose a simple yet efficient strategy for imaging dynamic scenes that requires less restrictive hypotheses. Our method relies on a hybrid device that combines a standard imaging arm with a hyperspectral single-pixel imaging arm. By estimating the motion from the standard arm, we show how to compensate the motion perceived by the single-pixel arm. In Section 2, we model dynamic SPI and static reconstruction. In Section 3, we present the proposed motion-compensated reconstruction strategy that we evaluate using the simulations and experimental measurements described in Section 4. We report and discuss our results in Section 5.

2. SINGLE-PIXEL IMAGING

2.1. Forward model

We model the single-pixel camera (SPC) acquisition as a series of scalar products between the unknown scene and some light patterns h_k, $1 \leq k \leq K$, that are sequentially loaded on a DMD. We consider a dynamic scene $f(\mathbf{x}, t)$ that varies

This work was supported by the French National Research Agency (ANR), under Grant ANR-22-CE19-0030-01 (ULHYB Project).

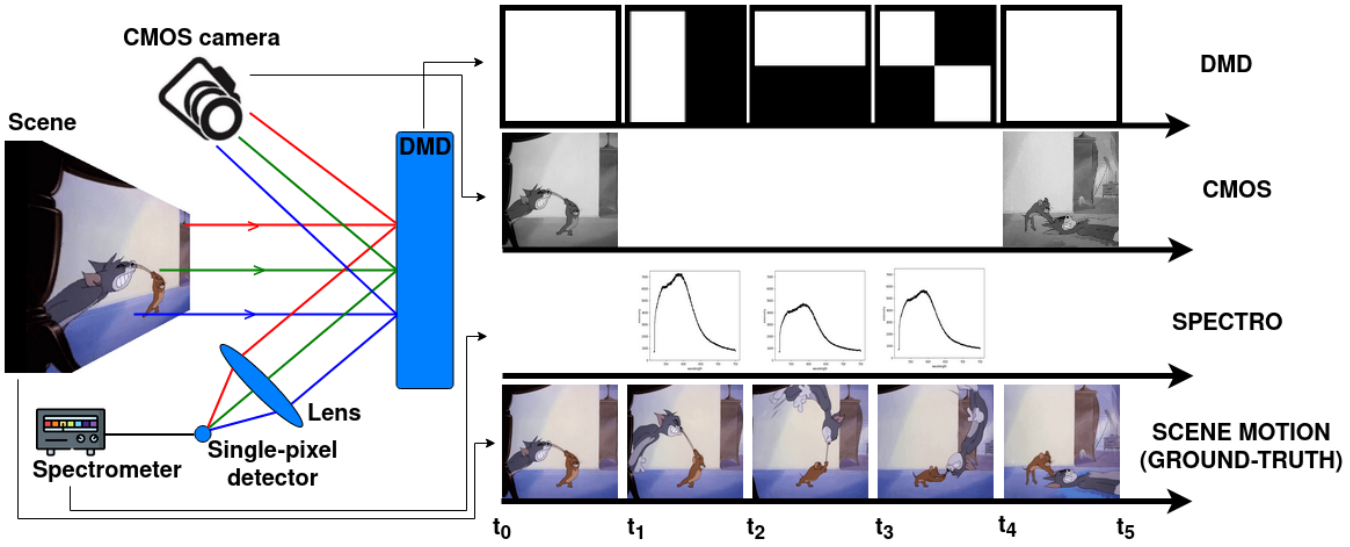

Fig. 1. Setup used for the acquisitions (left hand-side) and acquisition process between two CMOS frames (right hand-side). Light is emitted from the scene and projected onto the DMD. Depending on the orientation of the DMD, light rays are either directed towards the CMOS camera for motion estimation of the scene, or are summed on the single-pixel detector thanks to a converging lens. The single-pixel camera is then linked to a spectrometer for hyperspectral imaging.

during the acquisition. Assuming that the k-th DMD pattern is displayed during the time interval $[t_k, t_{k+1}]$, the k-th measurement (for all spectral channels) is given by

$$m_k = \int_X h_k(\mathbf{x}) f_k(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{1}$$

where $f_k(\mathbf{x}) = \int_{t_k}^{t_{k+1}} f(\mathbf{x}, t) \, \mathrm{d}t$ represents the k-th frame of the scene and X is the SPC field of view. Our goal is to reconstruct all the frames of the dynamic scene given a set of K measurements. This is a very challenging problem where K functions need to be estimated from K scalar measurements.

2.2. The static case

In the case of a static scene $f(\mathbf{x}, t) = f_{\text{ref}}(\mathbf{x})$, all frames are identical and the forward problem can be discretized as

$$\boldsymbol{m} = \boldsymbol{H} \boldsymbol{f}_{\text{ref}}, \tag{2}$$

where $\boldsymbol{m} \in \mathbb{R}^K$ is the measurement vector, $\boldsymbol{f}_{\text{ref}} \in \mathbb{R}^N$ is the discrete (unknown) image of N pixels and $\boldsymbol{H} \in \mathbb{R}^{K \times N}$ is the measurement matrix that contains the DMD patterns. In the presence of noise and/or when $K \leq N$, a good estimate of $\boldsymbol{f}_{\text{ref}}$ can be obtained by optimization of a hand-crafted objective

$$\min_{\boldsymbol{f}} \frac{\eta}{2} \|\boldsymbol{H}\boldsymbol{f} - \boldsymbol{m}\|_2^2 + \mathcal{R}(\boldsymbol{f}), \tag{3}$$

where η is a penalty parameter and $\mathcal{R}$ is a user-defined regularizer (e.g., $\|\cdot\|_2^2$ (L2) or total-variation (TV)). For high noise level and subsampling ratios, powerful alternatives include deep learning-based reconstruction [14, 15].

As shown in Section 5, however, the use of the static forward model given by (2) results in strong artifacts in the case of dynamic scenes.

3. PROPOSED DYNAMIC RECONSTRUCTION

3.1. Motion compensation

We assume that the scene can be motion-compensated, i.e.

$$f_k(v_k(\mathbf{z})) = f_{\text{ref}}(\mathbf{z}), \tag{4}$$

where $f_{\text{ref}} \in L^2(Z_k)$ represent a reference static image and $v_k \colon Z_k \to X$ represents the deformation field of the scene between the k-th frame and the reference frame.

Proposition. *Given $f_k(v_k(\mathbf{z})) = f_{\text{ref}}(\mathbf{z})$, where $v_k \colon Z_k \to X$ is bijective and $\operatorname{supp}(f_{\text{ref}}) = X$, the measurement vector can be written as*

$$\boldsymbol{m} = \boldsymbol{H}_{\text{dyn}} \boldsymbol{f}_{\text{ref}}, \tag{5}$$

where $\boldsymbol{H}_{\text{dyn}}^{K \times N}$ represents the dynamic measurement matrix and $\boldsymbol{f}_{\text{ref}} \in \mathbb{R}^N$ is the discrete reference image.

Proof. By substitution of (4) in (1), followed by the change of variable $\mathbf{x} = v_k(\mathbf{z})$, we obtain:

$$m_k = \int_X h_k(\mathbf{x}) f_{\text{ref}}(v_k^{-1}(\mathbf{x})) \, \mathrm{d}\mathbf{x} \tag{6}$$

$$= \int_{Z_k} h_k(v_k(\mathbf{z})) f_{\text{ref}}(\mathbf{z}) \, |\det J_k(\mathbf{z})| \, \mathrm{d}\mathbf{z} \tag{7}$$

$$= \int_X h_k^{\text{dyn}}(\mathbf{z}) f_{\text{ref}}(\mathbf{z}) \, \mathrm{d}\mathbf{z}, \tag{8}$$

where $h_k^{\text{dyn}}(\mathbf{z}) = h_k(v_k(\mathbf{z}))\,|\det J_k(\mathbf{z})|$, with J_k being the Jacobian of v_k. Discretization of (8) completes the proof. □

Provided that the deformation field v_k is known, the dynamic reconstruction problem simplifies to the resolution of a linear system as in the static case. In accordance with (7), we need to construct the dynamic matrix $\boldsymbol{H}_{\text{dyn}}$ from the static matrix $\boldsymbol{H}$ by applying the deformation field in a row-by-row manner. In practice, the dynamic matrix turns out to be ill-conditioned. This highlights the need to adopt a similar variational approach as in (3) to reconstruct the reference image.

3.2. Hybrid estimation of the deformation field

We propose a hybrid approach whereby the deformation field is estimated from the video stream acquired by a CMOS camera monitoring the same scene as the SPC. As shown in Fig. 1, this is achieved by uploading white patterns periodically onto the DMD such that the CMOS camera captures the entire scene. The CMOS camera and the DMD are synchronized so that each frame corresponds to the display of a white pattern.

The deformation field can be determined by optical-flow methods [16]. As the deformation field is known only every few patterns, we use linear interpolation to acquire the deformations for all frames. The obtained deformation field, denoted v_k^{cmos}, is given in the spatial coordinates of the CMOS camera. We retrieve the deformation field in the spatial coordinate of the single-pixel camera by assuming the existence of a mapping $\mathbf{x} = \mathcal{G}(\mathbf{x}^{\text{cmos}})$ that maps any point in the CMOS image to its corresponding position in the single-pixel image. For instance, we can choose $\mathcal{G}$ as a homography that is calibrated once in a preliminary step with a direct linear transform (DLT) [17]. By definition of the CMOS deformation field, we have $v_k^{\text{cmos}}(\mathbf{z}^{\text{cmos}}) = \mathbf{x}_k^{\text{cmos}}$. By exploiting the camera mapping, we obtain $v_k^{\text{cmos}}(\mathcal{G}^{-1}(\mathbf{z})) = \mathcal{G}^{-1}(\mathbf{x}_k)$ and, therefore, $\mathcal{G}(v_k^{\text{cmos}}(\mathcal{G}^{-1}(\mathbf{z}))) = \mathbf{x}_k$. We can finally identify the single-pixel deformation field as

$$v_k = \mathcal{G} \circ v_k^{\text{cmos}} \circ \mathcal{G}^{-1}. \tag{9}$$

4. EXPERIMENTS

4.1. Numerical simulations

We simulate the dynamic acquisition of a brain surface subject to a periodic affine motion model

$$v_k(\mathbf{x}) = c + \begin{pmatrix} s(t_k)^{-1} & 0 \\ 0 & s(t_k) \end{pmatrix} (\mathbf{x} - \mathbf{c}), \tag{10}$$

where $s(t) = a\sin(\frac{2\pi t}{T}) + 1$, a is the motion amplitude, T the motion period and $\mathbf{c}$ is the coordinate of the center of the image. We consider the acquisition of $K = 128^2$ Hadamard patterns and corrupt the measurements by Poisson noise assuming the maximum image intensity is 1000 photons. The penalty term is empirically set to $\eta = 10^{-3}$ and $\eta = 8 \cdot 10^{-3}$ for L2 and TV regularization. We consider the reconstruction of 128×128 images.

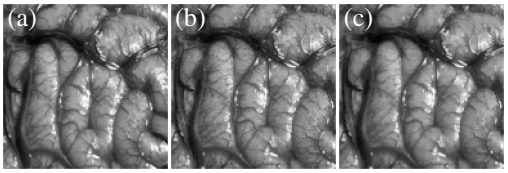

Fig. 2. Simulation results: reconstruction of a static brain surface. From left to right: the reference image, the static reconstruction (no regularization), the static TV reconstruction with $\eta = 8 \cdot 10^{-3}$.

4.2. Experimental acquisitions

The experimental setup is composed of a 1024×768 DMD (ViALUX GmbH DLP V-700), a spectrometer (Avantes AvaSpec-ULS2048CL-EVO) and an IDS CMOS camera (UI-3880CP-M-GL Rev 2). The CMOS camera has a resolution of 3088×2076 from which a 768×544 field of view is extracted. We image a diaphragm using a white LED (Thorlabs LIUCWHA) resulting in a bright disk. We consider two dynamic scenarios where i) the diameter of the diaphragm is fixed to 2 mm while its center is translated across the field of view and ii) the center of the diaphragm is fixed while its diameter varies between 1 mm and 10 mm. We consider the acquisition of $K = 64^2$ Hadamard patterns and the reconstruction of 64×64 images. The illumination time $\Delta t = t_{k+1} - t_k$ is set to 1.4 ms. The two datasets, in addition to many more, are available in the SPIHIM collection [2].

5. RESULTS AND DISCUSSION

5.1. Simulation study

Figure 2 depicts the reconstruction for a static scene, showcasing the effective handling of noise through TV regularization. Reconstruction results when the brain surface image is subject to the affine motion described in (10) are given in Fig. 3 for two motion amplitudes a. The first two rows are obtained loading the patterns in the natural order [18] while the third row used the order proposed in [19] that was found to improve the reconstruction quality for sub-sampled acquisitions. As expected, we observe stronger motion artifacts when reconstructing the brain surface using the static model (first column of Fig. 3). These artifacts cannot be corrected simply using TV regularization (second column of Fig. 3). However, one can see that our motion compensation method effectively allows the reconstruction of the reference image with the help of a L2 regularization to cope with the ill-posed nature of the problem (third column of Fig. 3). Small residual motion artifacts are still visible in Fig. 3g and 3k when motion is too strong. This can be explained by the interpolation of the

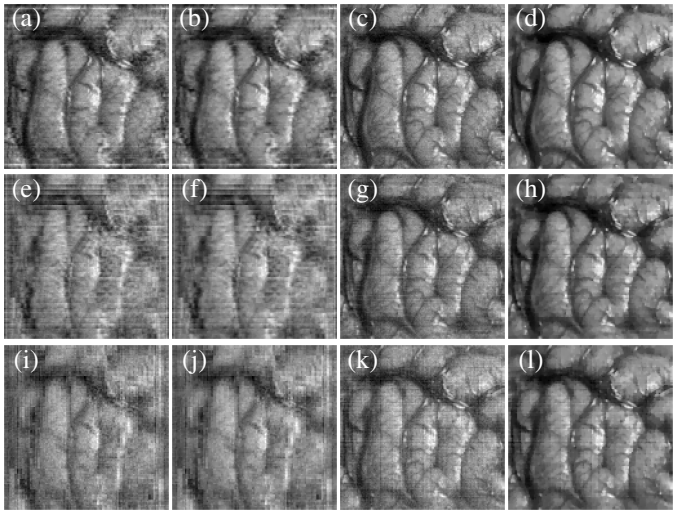

Fig. 3. Simulation results: brain surface subjected to the motion given in (10) and Poisson noise. Column-wise, the figure represents the static (no regularization), the static-TV, the dynamic-L2 and the dynamic-TV reconstructions. (a)-(d) consider a motion amplitude $a = 0.05$, for (e)-(l) $a = 0.2$. The measurements for (a)-(h) were acquired by loading the patterns in the natural order while (i)-(l) used the optimized order. The penalty term is set to 10^{-3} and $8 \cdot 10^{-3}$ for L2 and TV regularization.

deformation field which leads to systematic errors. We can attenuate these artifacts using the motion-compensated method with TV regularization (fourth column of Fig. 3). We note that even though TV regularization is efficient for noise removal, it can erase small blood vessels in the reconstructed brain. Concerning the order for loading the patterns, blur artifacts were reduced for the static reconstructions using the optimized order, as can be seen in Fig. 3i and 3j. Dynamic reconstruction of the scene is still achieved in Fig. 3k and 3l even though the residual artifacts are different.

Table 1 reports quantitative metrics - peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) - for the different reconstruction methods. These metrics confirm the visual evaluation. The combination of the proposed motion compensation method and TV regularization outperforms all other methods for dynamic scenes. The smaller the motion amplitude, the closer the scores approach those of the static brain surface. The advantage of TV over L2-regularization is illustrated by a systematic enhancement of both PSNR and SSIM. Concerning the pattern loading order, the optimized order enhanced PSNR scores for static and dynamic-TV reconstructions, supporting its use for experimental acquisitions.

5.2. Experimental validation

Reconstructions from experimental data are displayed in Fig. 4. The superposed contour of the conventional reference image shows the importance of dynamic reconstructions over static reconstructions. Strong artifacts are present when the motion is not taken into account (first column), which cannot be removed with regularization (second column). On the contrary, the proposed motion compensation strategy allows the blur artifacts to be removed and a clean disk to be recovered as expected (last two columns of Fig. 4).

		Static	Static TV	Dynamic L2	Dynamic TV
$a = 0$	PSNR	32.57	**34.68**		
(nat. order [18])	SSIM	0.94	**0.96**		
$a = 0.05$	PSNR	20.16	20.39	27.77	**30.69**
(nat. order [18])	SSIM	0.54	0.57	0.87	**0.91**
$a = 0.1$	PSNR	18.05	18.20	26.27	**28.43**
(nat. order [18])	SSIM	0.39	0.41	0.84	**0.89**
$a = 0.2$	PSNR	16.12	16.20	24.39	**25.32**
(nat. order [18])	SSIM	0.27	0.28	0.81	**0.86**
$a = 0.2$	PSNR	17.92	18.09	22.40	**26.45**
(opt. order [19])	SSIM	0.30	0.32	0.74	**0.83**

Table 1. Simulation results: PSNR (dB) and SSIM scores of different reconstruction methods for several motion amplitudes and different pattern ordering.

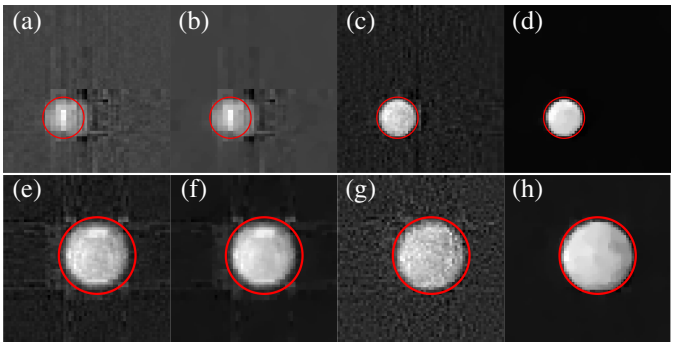

Fig. 4. Experimental results: white spot subjected to a rigid motion (first row) and a zooming motion (second row). The deformation was estimated with a CMOS camera using an optical-flow algorithm. A handmade contour of the CMOS reference frame is superposed in red to the static (no regularization), the static-TV, the dynamic-L2 and the dynamic-TV reconstructions.

6. CONCLUSION

We demonstrate, using numerical simulations and experimental acquisitions, an effective framework for dynamic single-pixel imaging based on a hybrid approach which exploits the images of the scene provided by a conventional camera. The proposed method can be applied to all channels of the hyperspectral arm. A limitation of this work is that the use of TV regularization tends to erase small details, in particular when used on low spatial resolution images. In the future, we will use the high-resolution images from the standard camera to increase the spatial resolution of the SPC using pansharpening techniques.

7. COMPLIANCE WITH ETHICAL STANDARDS

This is a study for which no ethical approval was required.

8. REFERENCES

[1] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[2] Guilherme Beneti Martins, Laurent Mahieu-Williame, Thomas Baudier, and Nicolas Ducros, "OpenSpyrit: an ecosystem for open single-pixel hyperspectral imaging," *Optics Express*, vol. 31, no. 10, pp. 15599, May 2023.

[3] Guolan Lu and Baowei Fei, "Medical hyperspectral imaging: a review," *Journal of biomedical optics*, vol. 19, no. 1, pp. 010901–010901, 2014.

[4] Jaime J Bravo, Jonathan D Olson, Scott C Davis, David W Roberts, Keith D Paulsen, and Stephen C Kanick, "Hyperspectral data processing improves ppix contrast during fluorescence guided surgery of human brain tumors," *Scientific reports*, vol. 7, no. 1, pp. 9455, 2017.

[5] Matthew P Edgar, Graham M Gibson, and Miles J Padgett, "Principles and prospects for single-pixel imaging," *Nature photonics*, vol. 13, no. 1, pp. 13–20, 2019.

[6] Cian M Scannell, Adriana DM Villa, Jack Lee, Marcel Breeuwer, and Amedeo Chiribiri, "Robust non-rigid motion compensation of free-breathing myocardial perfusion mri data," *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1812–1820, 2019.

[7] Simon Rit, David Sarrut, and Laurent Desbat, "Comparison of analytic and algebraic methods for motion-compensated cone-beam ct reconstruction of the thorax," *IEEE transactions on medical imaging*, vol. 28, no. 10, pp. 1513–1525, 2009.

[8] Jieqing Jiao, Alexandre Bousse, Kris Thielemans, Ninon Burgos, Philip SJ Weston, Jonathan M Schott, David Atkinson, Simon R Arridge, Brian F Hutton, Pawel Markiewicz, et al., "Direct parametric reconstruction with joint motion estimation/correction for dynamic brain pet data," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 203–213, 2016.

[9] Aswin C Sankaranarayanan, Lina Xu, Christoph Studer, Yun Li, Kevin F Kelly, and Richard G Baraniuk, "Video compressive sensing for spatial multiplexing cameras using motion-flow models," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 1489–1518, 2015.

[10] Shuming Jiao, Mingjie Sun, Yang Gao, Ting Lei, Zhenwei Xie, and Xiaocong Yuan, "Motion estimation and quality enhancement for a single image in dynamic single-pixel imaging," *Optics Express*, vol. 27, no. 9, pp. 12841, Apr. 2019.

[11] Sagi Monin, Evgeny Hahamovich, and Amir Rosenthal, "Single-pixel imaging of dynamic objects using multi-frame motion estimation," *Scientific Reports*, vol. 11, no. 1, pp. 7712, Apr. 2021, Number: 1 Publisher: Nature Publishing Group.

[12] Zijun Guo, Wenwen Meng, Dongfeng Shi, Linbin Zha, Wei Yang, Jian Huang, Yafeng Chen, and Yingjian Wang, "Fast localization and single-pixel imaging of the moving object using time-division multiplexing," *arXiv preprint arXiv:2208.07371*, 2022.

[13] Antonio Lorente Mur, Françoise Peyrin, and Nicolas Ducros, "Recurrent neural networks for compressive video reconstruction," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 1651–1654.

[14] Antonio Lorente Mur, Pierre Leclerc, Françoise Peyrin, and Nicolas Ducros, "Single-pixel image reconstruction from experimental data using neural networks," *Optics Express*, vol. 29, no. 11, pp. 17097–17110, 2021.

[15] Antonio Lorente Mur, Françoise Peyrin, and Nicolas Ducros, "Deep expectation-maximization for single-pixel image reconstruction with signal-dependent noise," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 759–769, 2022.

[16] Christopher Zach, Thomas Pock, and Horst Bischof, "A duality based approach for realtime tv-l 1 optical flow," in *Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29*. Springer, 2007, pp. 214–223.

[17] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[18] William K Pratt, Julius Kane, and Harry C Andrews, "Hadamard transform image coding," *Proceedings of the IEEE*, vol. 57, no. 1, pp. 58–68, 1969.

[19] Catherine F Higham, Roderick Murray-Smith, Miles J Padgett, and Matthew P Edgar, "Deep learning for real-time single-pixel video," *Scientific reports*, vol. 8, no. 1, pp. 2369, 2018.