Deep neural networks for single-pixel compressive video reconstruction

Antonio Lorente Mur, Bruno Montcel, Françoise Peyrin, Nicolas Ducros

Deep neural networks for single-pixel compressive video reconstruction

Antonio Lorente Mur[a], Bruno Montcel[a], Françoise Peyrin[a], and Nicolas Ducros[a]

[a]Univ Lyon, INSA-Lyon, UCB Lyon 1,CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France

ABSTRACT

Single-pixel imaging is a paradigm that enables the capture of an image from a single point detector using a spatial light modulator. This approach is particularly interesting for optical set-ups where pixelated arrays of detectors are either too expensive or too cumbersome (*e.g.*, multispectral, infrared imaging). It acquires the inner product between the image of the scene and a set of user-defined patterns that are sequentially uploaded onto the spatial light modulator. Compressed data acquisition reduces the acquisition time, although it leads to an ill-posed reconstruction problem, which is very challenging for real-time applications.

Recently, neural networks have emerged as competitive alternatives to traditional reconstruction methods. Neural networks are parametric models that are trained by exploiting large datasets. Their noniterative nature allows for fast reconstructions, which opens the door to real-time image reconstruction from compressed acquisition.

In this study, we evaluate the different networks for static and dynamic imaging. In particular, we introduce a recurrent neural network that is designed to exploit the spatiotemporal redundancy in videos via a memory state. We validate our algorithms on simulated data from the UCF-101 dataset, with a resolution of 128×128 pixels and a compression ratio of 98%. We also show experimentally that we can resolve small spectral differences in the spectrum of human skin measured *in vivo*.

Keywords: Hyperspectral imaging, deep learning, image reconstruction, video reconstruction, computational optics, single-pixel camera

1. INTRODUCTION

A single-pixel (SP) camera is a compressive imager that uses a single point detector to recover a two-dimensional image.[1] With a set of lenses and a digital micro-mirror device (DMD), the set-up can acquire the inner product between the scene and some user-defined light patterns. SP imaging has been used successfully for fluorescence microscopy,[2] hyperspectral imaging,[3,4] diffuse tomography,[5] and image-guided surgery.[6] As SP measurements are performed sequentially, it is necessary to limit the number of light patterns for real-time applications. Therefore, the reconstruction problem of SP imaging is typically an under-determined inverse problem.

Under-determination is classically addressed using ℓ_2-[7] or ℓ_1- (or total variation)[1] regularization. On the one hand, ℓ_2 approaches are fast, but they can lead to reduced image quality. On the other hand, while ℓ_1 approaches lead to improved image quality, they require time-consuming iterative algorithms. Recently, deep neural networks have shown promising results for image reconstruction.[8,9] 10 proposed an auto-encoder network for SP image acquisition and reconstruction. This represents a useful step towards real-time imaging, although the use of a fully connected layer with more than 99.5% of the network parameters is not well understood yet. Moreover, this approach processes each frame independently, which fails to exploit the spatio-temporal redundancy in video sequences.

In this study, we first interpret the SP reconstruction problem as a completion problem where the missing measurements have to be estimated. Given an image database, we derive analytically the best linear solution of the completion problem. Then, we freeze the fully connected layer of a network so as to implement this solution, and train only the convolution layers downstream. Freezing the fully connected layer significantly decreases the number of parameters to be learnt. Finally, we then propose a fast deep-learning reconstructor that exploits the spatio-temporal features in a video. In particular, we consider a recurrent neural network (RNN) that is suited to handling image sequences through its internal state that memorizes previous inputs. Among RNNs, the long short-term memory cells are probably the most popular deep-learning

Further author information: (Send correspondence to Nicolas Ducros)
Nicolas Ducros : E-mail: nicolas.ducros@creatis.insa-lyon.fr

variant.[11] Here, we consider gated recurrent units, which have been shown to have similar performance to long short-term memory cells,[12] although they have less memory requirement.

This paper is organized as follows. In section 2, we introduce the mathematical framework of SP imaging, alongside the classic SP reconstruction approaches. In Section 3, we first give an interpretation of the use of a fully connected layer to project the raw measurement data into the image domain. We also describe our proposed RNN for solving the SP video problem. Section 4 reports our numerical simulations and experimental results, which are discussed further in Section 5.

2. SINGLE-PIXEL VIDEO RECONSTRUCTION

2.1 Single-pixel acquisition

Let $(\boldsymbol{f}_t)_{t\in\mathbb{N}} \in \mathbb{R}^{N\times 1}$ be a video sequence, where $\boldsymbol{f}_t$ is the t-th frame in the video sequence. With a SP camera, we have access to a measurement sequence $(\boldsymbol{m}_t)_{t\in\mathbb{N}} \in \mathbb{R}^{K\times 1}$, as given by[1]

$$\boldsymbol{m}_t = \boldsymbol{P}\boldsymbol{f}_t \Delta t, \quad \forall t, \tag{1}$$

where $\boldsymbol{P} \in \mathbb{R}^{M\times N}$ is the sequence of patterns that are uploaded onto the DMD, and Δt is the integration time. For each time frame, $\boldsymbol{P} = (\boldsymbol{p}_{t,1}, \ldots, \boldsymbol{p}_{t,M})^\top \in \mathbb{R}_+^{M\times N}$ is a matrix that contains a sequence of M patterns. These patterns can be chosen on an orthogonal basis (*e.g.*, Hadamard;[13] Fourier;[14] wavelets[7]). There are several ways to implement patterns with negative values (see 15 for details). Here, we consider the splitting methods where each pattern is separated into its positive and negative components. We finally assume that $\boldsymbol{f}_t$ is slowly varying over a period of time $M\Delta t$, which corresponds to the acquisition of each measurement vector $\boldsymbol{m}_t$.

2.2 Static single-pixel reconstruction

Static reconstruction recovers $\boldsymbol{f}_t^* \approx \boldsymbol{f}_t$ by designing reconstruction schemes Φ that rely solely on the current measurements; *i.e.*, $\boldsymbol{f}_t^* = \Phi(\boldsymbol{m}_t), \forall t$. Traditional static approaches solve a sequence of optimization problems of the form

$$\boldsymbol{f}_t^* \in \operatorname{argmin} \mathcal{R}(\boldsymbol{f}_t) \quad \text{s.t.} \quad \boldsymbol{P}\boldsymbol{f}_t\Delta t = \boldsymbol{m}_t, \quad \forall t, \tag{2}$$

where $\mathcal{R}$ is typically the ℓ_2 norm[7]), the ℓ_1 norm,[1] or the total variation semi-norm.[16] To speed-up the reconstruction, 10 proposed the use of an auto-encoder network Φ, such that

$$\boldsymbol{f}_t^* = \Phi_{\boldsymbol{\theta}^*}(\boldsymbol{m}_t), \quad \forall t, \tag{3}$$

where $\boldsymbol{\theta}^*$ represents the weights of the network that are optimized during the training phase. Although the training phase is time consuming, the evaluation of Equation (3) is fast. However, this approach fails to exploit the spatio-temporal redundancy within the video sequences, as the same network $\Phi_{\boldsymbol{\theta}^*}$ is used for all of the time frames, and it has no feedback mechanism.

2.3 Dynamic reconstruction

Dynamic reconstructions can exploit temporal features by designing reconstruction operators Φ that take into account the measurements of the previous frames $(\boldsymbol{m}_{t'})_{0\leq t'\leq t}$ for the reconstruction of the current frame $\boldsymbol{f}_t$:

$$\boldsymbol{f}_t^* = \Phi(\boldsymbol{m}_t, \ldots, \boldsymbol{m}_0). \tag{4}$$

In particular, different studies that have proposed sparsity-promoting solutions rely on minimizing a problem of the form[17]

$$\boldsymbol{f}_t^* \in \operatorname{argmin} \mathcal{R}(\boldsymbol{f}_t, \ldots, \boldsymbol{f}_0) \quad \text{s.t.} \quad \boldsymbol{P}_{t'}\boldsymbol{f}_{t'}\Delta t' = \boldsymbol{m}_{t'}, \quad 0 \leq t' \leq t \tag{5}$$

However, such approaches require iterative schemes that lead to reconstruction times ($\sim$min) that are too long for real-time applications.

3. PROPOSED APPROACH

3.1 Static reconstruction through Bayesian completion

Let $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ be an image basis that includes the acquired patterns; *i.e.*, $\boldsymbol{Q} = [\boldsymbol{P}^\top, \boldsymbol{H}^\top]^\top$, where $\boldsymbol{H} \in \mathbb{R}^{L\times N}$, $L = N - M$, represents the missing patterns. On the assumption that $\boldsymbol{Q}$ is an orthogonal matrix, the traditional least-squares solution is given by

$$\boldsymbol{f}_t^* = \boldsymbol{Q}^\top \boldsymbol{y}_t^*, \quad \text{with } \boldsymbol{y}_t^* = \begin{bmatrix} \boldsymbol{m}_t \\ \boldsymbol{0} \end{bmatrix}, \tag{6}$$

The zero entries of $\boldsymbol{y}_t^*$ that correspond to coefficients that are not acquired can be estimated through their correlation with the acquired coefficients $\boldsymbol{m}_t$, by exploiting a database of natural images. Assuming that the measurement vector $\boldsymbol{m}_t$ is a sample of a Gaussian random vector, we have[18]

$$\boldsymbol{f}_t^* = \boldsymbol{Q}^\top \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_1^{-1} \end{bmatrix} \boldsymbol{m}_t + \boldsymbol{Q}^\top \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \end{bmatrix}, \tag{7}$$

where $\boldsymbol{\Sigma}_{21} \in \mathbb{R}^{L\times M}$ is the covariance between the missing and the acquired coefficients, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{M\times M}$ is the covariance matrix of the acquired coefficients, $\boldsymbol{\mu}_2$ is the mean of the measured coefficients, and $\boldsymbol{\mu}_2$ is the mean of the missing coefficients. We assume that the covariance matrix $\boldsymbol{\Sigma}_1$ is invertible.

Given an image database $\{\boldsymbol{f}^{(k)}\}_{k=1}^K$, we compute the previous statistics as

$$\boldsymbol{\mu}_1 = \frac{1}{K}\sum_k \boldsymbol{P}\boldsymbol{f}^{(k)}, \qquad \boldsymbol{\mu}_2 = \frac{1}{K}\sum_k \boldsymbol{H}\boldsymbol{f}^{(k)}, \tag{8}$$

$$\boldsymbol{\Sigma}_1 = \frac{1}{K-1}\sum_k (\boldsymbol{P}\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_1)(\boldsymbol{P}\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_1)^\top, \qquad \boldsymbol{\Sigma}_{21} = \frac{1}{K-1}\sum_k (\boldsymbol{H}\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_2)(\boldsymbol{P}\boldsymbol{f}^{(k)} - \boldsymbol{\mu}_1)^\top. \tag{9}$$

3.2 Static reconstruction based on deep learning

When the measurement vector $\boldsymbol{m}_t$ cannot be assumed to be a sample of a Gaussian random vector, the solution to the completion problem is nonlinear. We propose to learn this through a family of nonlinear mapping $\mathcal{H}_{\boldsymbol{\theta}}$ parameterized by $\boldsymbol{\theta}$. We consider a neural network model of the form

$$\mathcal{H}_{\boldsymbol{\theta}} = \mathcal{H}_{\boldsymbol{\theta}_L} \circ \ldots \circ \mathcal{H}_{\boldsymbol{\theta}_1} \tag{10}$$

where $\mathcal{H}_{\boldsymbol{\theta}_L}$, $1 \le \ell \le L$, is the ℓ-th layer of the network, and $\circ$ is the function composition.

The first layer is traditionally a fully connected layer that maps the measurement $\boldsymbol{m}_t \in \mathbb{R}^M$ to a raw solution $\tilde{\boldsymbol{f}}_t \in \mathbb{R}^N$, as shown in Fig. 1. Mathematically, the output of this layer can be expressed by

$$\tilde{\boldsymbol{f}}_t = \mathcal{H}_{\boldsymbol{\theta}_1}(\boldsymbol{m}_t) = \boldsymbol{W}\boldsymbol{m}_t + \boldsymbol{b} \tag{11}$$

Here, as in 18, we set the weights $\boldsymbol{W}$ and the biases $\boldsymbol{b}$ to provide the solution given by Equation (7). The rest of the layers of the network are trained considering the mean square error loss function

$$\boldsymbol{\theta}^* \in \underset{(\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_L)}{\arg\min} \frac{1}{K}\sum_{k=1}^K \|\mathcal{H}_{\boldsymbol{\theta}}(\boldsymbol{m}^{(k)}) - \boldsymbol{f}^{(k)}\|^2 \tag{12}$$

where $\{\boldsymbol{m}^{(k)} = \boldsymbol{P}_1\boldsymbol{f}^{(k)}\Delta t\}_{k=1}^K$ are the measurements associated to the image database $\{\boldsymbol{f}^{(k)}\}_{k=1}^K$.

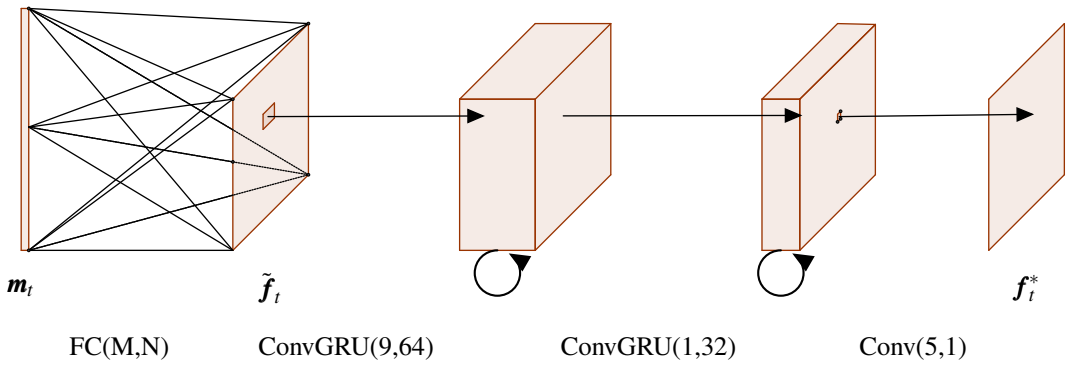

Figure 1: Proposed recurrent neural network for single-pixel video reconstruction. The design is inspired by 10, with the convolutional layers replaced by convolutional gated recurrent units (ConvGRUs) that maintain the long-term temporal dependency, as in 19. ConvGRU(K_s, F_m) designates a ConvGRU cell with convolutional kernels of size $K_s \times K_s$ and F_m output feature maps.

3.3 Dynamic reconstruction using recurrent neural networks

To exploit the temporal redundancy of video sequences, we use recurrent layers that have a hidden memory state. The current frame $\boldsymbol{f}_t^*$ is estimated from the current measurement vector $\boldsymbol{m}_t$ and the previous hidden state $\boldsymbol{h}_{t-1}$

$$(\boldsymbol{f}_t^*, \boldsymbol{h}_t) = \mathcal{G}_{\boldsymbol{\theta}}(\boldsymbol{m}_t, \boldsymbol{h}_{t-1}) \tag{13}$$

where $\mathcal{G}$ represents the RNN and $\boldsymbol{\theta}$ are the parameters of the RNN.[19] Note that the hidden state is also updated, so as to maintain long-term dependency.

We propose the four-layer network depicted in Fig. 1. The first layer is a fully connected layer that projects the measurement frame $\boldsymbol{m}_t$ to the image domain, as explained in Section 3.2. The next two layers are two convolutional gated recurrent units,[20] which are followed by a (regular) convolutional layer. Given a video database $\{\boldsymbol{f}_t^{(k)}\}_{1\le k\le K, 1\le t\le T}$, where $\boldsymbol{f}_t^{(k)} \in \mathbb{R}^N$ is the t-th frame in the k-th video, we trained our network as

$$\boldsymbol{\theta}^* \in \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{t=1}^{T} \frac{\|\boldsymbol{f}_t^{(k)} - \mathcal{G}_{\boldsymbol{\theta}}(\boldsymbol{m}_t^{(k)}, \boldsymbol{h}_{t-1}^{(k)})\|_2^2}{2ST} + \lambda \|\boldsymbol{\theta}\|_2^2 \tag{14}$$

where $\boldsymbol{m}_t^{(k)} = \boldsymbol{P}\boldsymbol{f}_t^{(k)}\Delta t$ is the measurement vector for t-th frame in the k-th video, and λ is the weight decay parameter. Weight decay is a classical regularization procedure that ensures the convergence of the training.

4. RESULTS AND DISCUSSION

4.1 Numerical experiments

We chose $M = 333$ Hadamard patterns of size $N = 128 \times 128$. Our choice for M is mostly motivated by the implementation of a real-time algorithm. With a DMD refresh rate of 20 kHz, and splitting the Hadamard patterns into positive and negative patterns,[15] the choice of $M = 333$ leads to a frame rate of about 30 frames per second.

The size of the convolutional kernels and the number of feature maps of our RNN are chosen to mimic those in 10. We trained and tested our network using the UCF-101[21] dataset. This is an action recognition dataset that consists of 13 320 videos from 101 action categories. We down-sampled all of the frames to 128×128 pixels.

The recurrent network is implemented using Pytorch[22] and trained using the ADAM[23] optimiser for 60 epochs. The step size is initialized to 10^{-3}, and divided by 5 every 40 epochs. Each sample consists of 10 consecutive frames that are randomly sampled from each video. We set the weight decay parameter λ to 10^{-6}. The number of learned parameters is

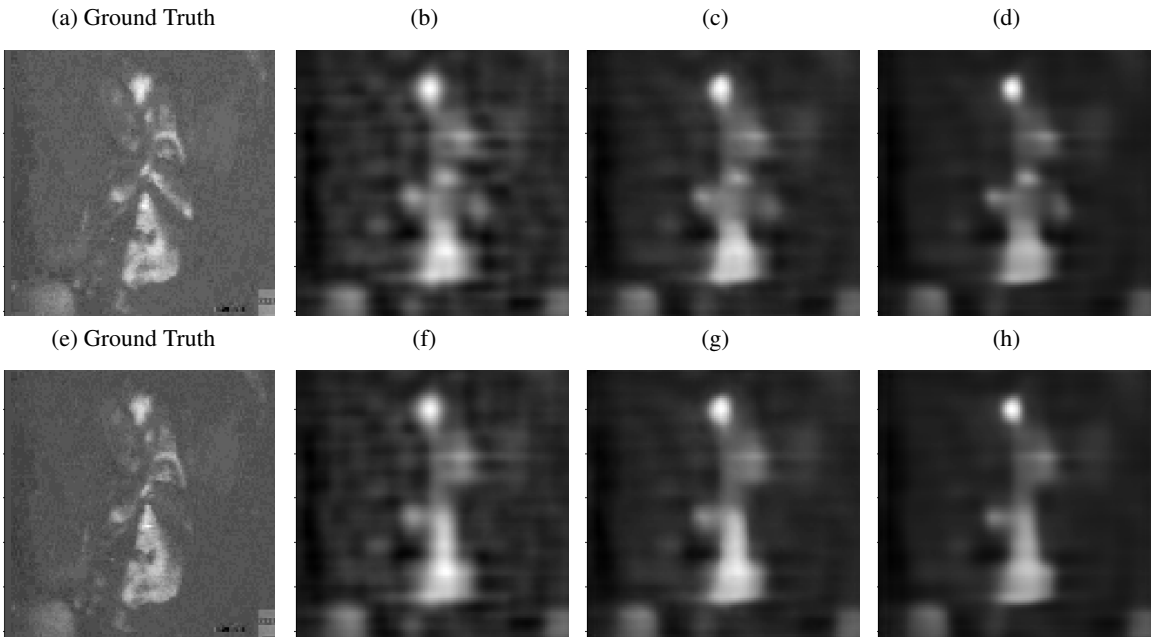

Figure 2: Reconstruction of two frames of a fluorescence-guided neurosurgery video sequence. (a) Ground truth of frame #10. (b) Completion method: peak signal-to-noise ratio (PSNR) = 26.76; structural similarity (SSIM) = 0.83. (c) Static network proposed in 10: PSNR = 26.85; SSIM = 0.83; (d) Proposed recurrent network: PSNR = 27.09; SSIM = 0.84. (e) Ground truth of frame #15. (f) Completion method: PSNR = 27.83; SSIM = 0.85. (g) Static network proposed in 10: PSNR = 27.96; SSIM = 0.85. (h) Proposed recurrent network: PSNR = 28.17; SSIM = 0.86.

1 021 185. Note that the fully connected layer is computed beforehand, as explained in Section 3, so our network does not learn it.

Table 1: Average peak signal-to-noise ratio (PSNR) and average structural similarity (SSIM) over the UCF-101 test dataset for the three different methods, as the least-squares solution (see Equation (6)), statistical completion (section 3.1), and the static network,[10] plus our proposed recurrent network (section 3.3).

Method	PSNR	SSIM
Least-squares solution (6)	20.81	0.9013
Completion method (section 3.1)	21.77	0.9205
Static network[10]	22.17	0.9255
Proposed recurrent network (section 3.3)	**22.25**	**0.9263**

In Table 1, we compare the average peak signal-to-noise ratio (PSNR) and the average structural similarity (SSIM) of the output of our network with the output of the network of 10, and with the result of the statistical completion (section 3.1) and that of the ℓ_2 regularisation, or the least-squares solution of Equation (6). We compute the PSNR and SSIM over the whole video frame-by-frame, and we average the frame-wise PSNR and SSIM over every video. We note that the recurrent network outperforms the other methods, and it is followed closely by the static network. We also note that despite being outperformed by the two networks, the completion method yields much higher scores than the pseudo-inverse method.

We display some sample frames of a video from the testing set in Fig. 2. It appears that indeed the estimate for the recurrent network shows greater accuracy compared to the other methods.

4.2 Experimental measurements

To validate our reconstruction methods, we also built the experimental set-up depicted in Fig. 3. The telecentric lens (Edmund Optics 62901) is positioned such that its image side projects the image of the scene onto the DMD (vialux V-7001), which is positioned at the object side of the lens. The DMD can implement different light patterns by reflection

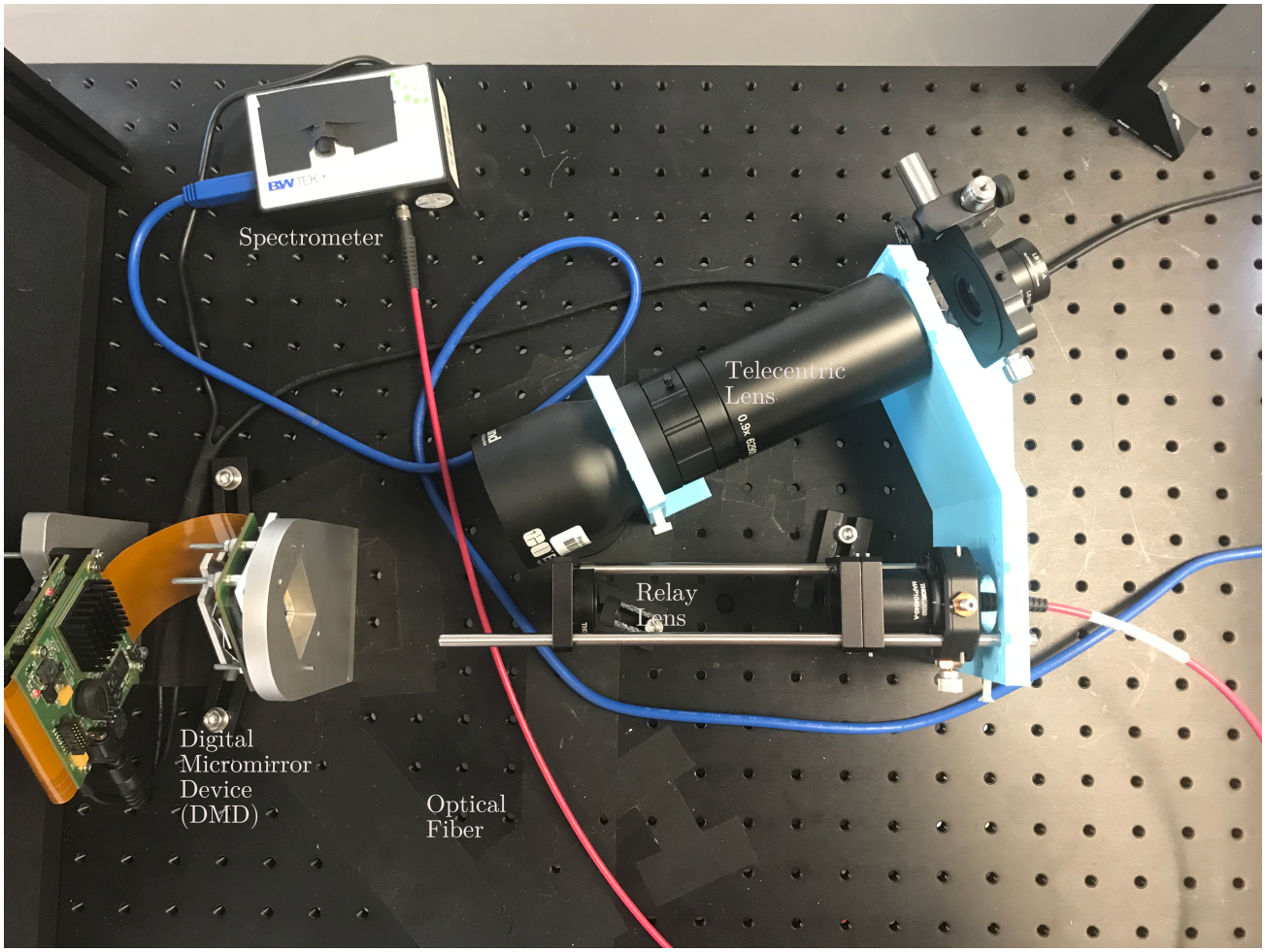

Figure 3: Optical set-up of the single-pixel camera.

of the incident light onto a relay lens, which injects the light into an optical fiber (Thorlabs FT1500UMT 0.39NA). This optical fiber is connected to a spectrometer (BWTek examplar model N^o BRC115P-V-ST1). We sequentially upload $M = 333$ patterns onto the DMD, each of dimension $N = 64 \times 64$, which gives a compression ratio of 92%. One spectrum is acquired by the spectrometer for each pattern over 15 ms, to give the total acquisition time of 1.3 s. As depicted in Fig. 4, we image the back of the hand of a human subject, which shows intact skin, a keloidal scar, and superficial vein regions. Then, we reconstruct a 64×64 image with the statistical completion method, using the open-source SPIRiT Matlab Toolbox.[24] The results are shown in Fig. 5.

We can identify the three regions of interest here: the keloidal scar, a vein, and the regular skin. Therefore, we extract some pixels in each of these regions (see Fig. 4) and compute the average spectrum (see Fig. 6). Each spectrum is normalized by the spectrum of the halogen lamp. The resulting spectra are in agreement with the biomedical literature. We observe the blue color of the vein, which has the same amount of blue as for the regular skin, but less intensity in the red wavelengths.[25] The red keloidal scar is less blue than the rest of the skin, although it has many high components in the red wavelengths.[26] We finally integrate the three spectra below 600 nm (blue region) and above 600 nm (red region) to quantify the blue-to-red ratio of the spectra (see Tab.2). We observe the lowest ratio for the keloidal scar (less blue than the skin) and the highest for the vein (less red than the skin), as expected.

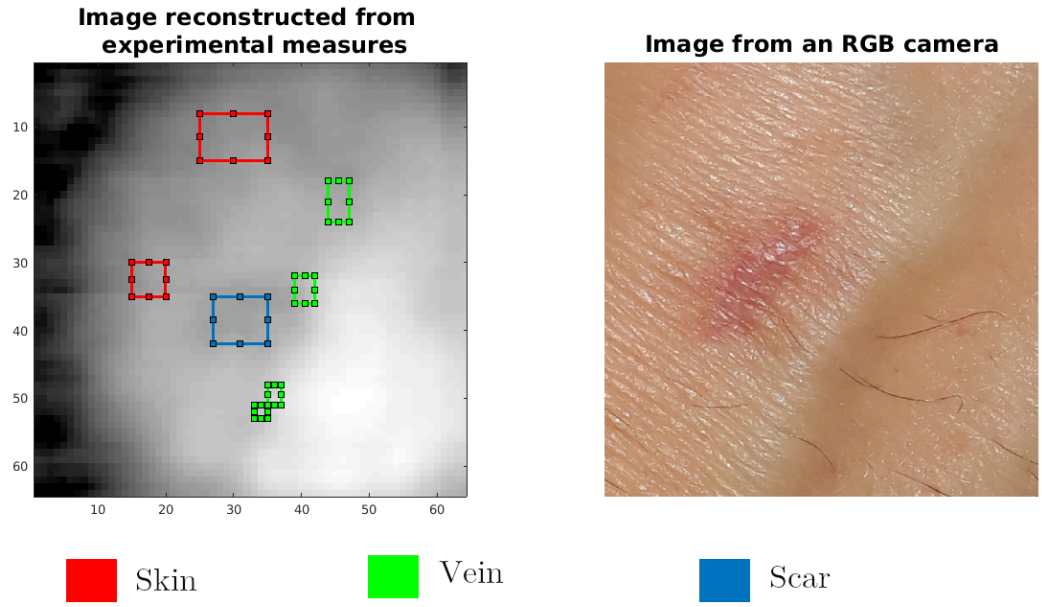

Figure 4: Back of a hand imaged by an RGB camera (right) and our compressive hyperspectral camera (left). The three regions of interest are superimposed on the gray-scale compressive image.

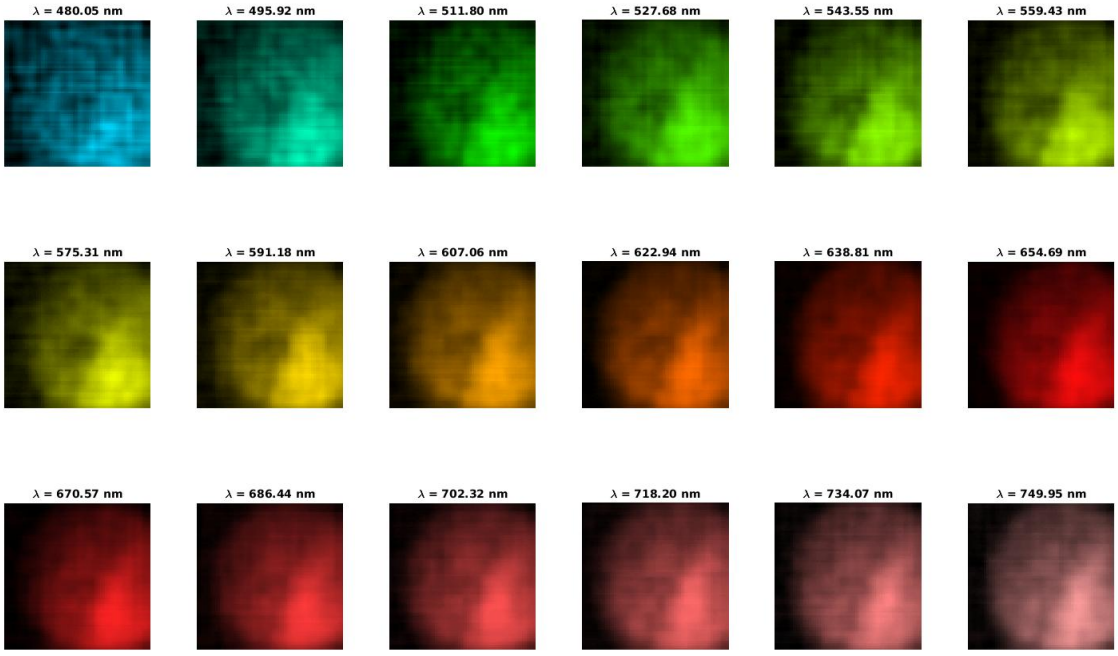

Figure 5: Hyperspectral reconstruction of the back of a hand, with light from a halogen lamp.

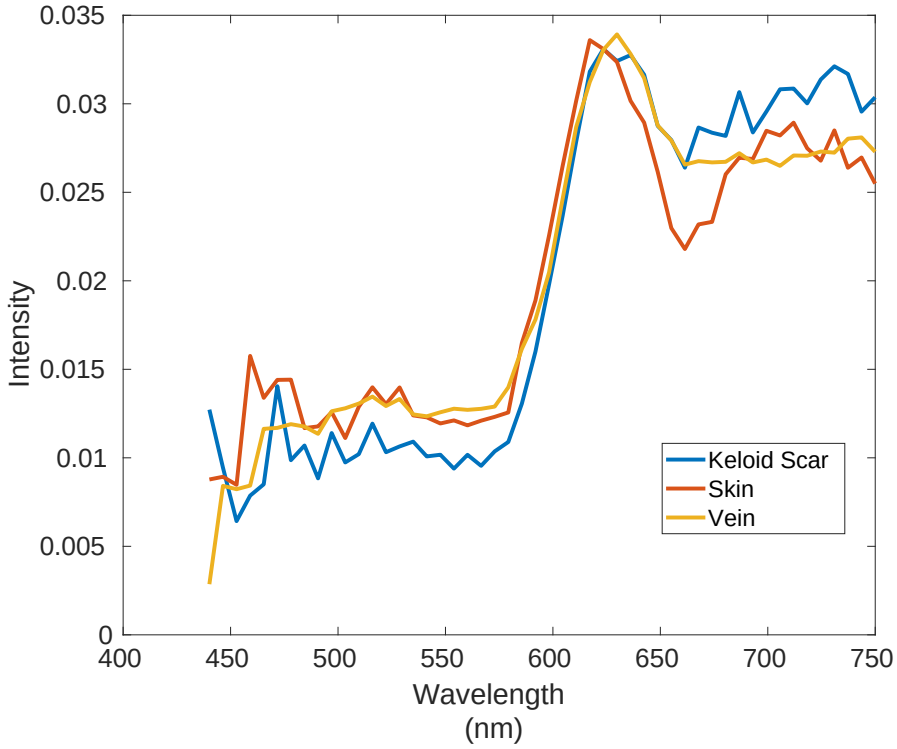

Figure 6: Average spectra in the three regions of interest (skin, vein, keloidal scar).

Table 2: Average color values for the different regions of interest

Region of interest	$\lambda < 600$ nm	$\lambda > 600$ nm
Keloidal scar	0.283	0.717
Skin	0.321	0.679
Vein	0.341	0.659

5. CONCLUSIONS

We demonstrate the interest for recurrent networks for video reconstruction using numerical simulations. We also show experimentally the interest for our Bayesian completion reconstruction method for hyperspectral data. In particular, we demonstrate that we can resolve small spectral differences in the spectrum of human skin measured *in vivo*.

In future work, we aim to take into account measurement noise and to evaluate our recurrent network using experimental data. We also plan to improve our dynamic reconstruction by designing other recurrent network architectures.

ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (ANR), under Grant ANR-17-CE19-0003 (ARMONI Project). It was performed within the framework of the LABEX PRIMES (ANR-11-LABX-0063) of the Université de Lyon, within the programme "Investissements d'Avenir" (ANR-11-IDEX-0007), operated by the ANR.

REFERENCES

[1] Duarte, M., Davenport, M., Takhar, D., Laska, J., Sun, T., Kelly, K., and Baraniuk, R., "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE* **25**, 83–91 (March 2008).

[2] Studer, V., Bobin, J., Chahid, M., Mousavi, H. S., Candes, E., and Dahan, M., "Compressive fluorescence microscopy for biological and hyperspectral imaging," *Proceedings of the National Academy of Sciences* **109**(26), E1679–E1687 (2012).

[3] Rousset, F., Ducros, N., Peyrin, F., Valentini, G., D'Andrea, C., and Farina, A., "Time-resolved multispectral imaging based on an adaptive single-pixel camera," *Opt. Express* **26**, 10550–10558 (Apr 2018).

[4] Arce, G. R., Brady, D. J., Carin, L., Arguello, H., and Kittle, D. S., "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine* **31**, 105–115 (Jan 2014).

[5] Pian, Q., Yao, R., Sinsuebphon, N., and Intes, X., "Hyperspectral compressive single-pixel imager for fluorescence lifetime sensing," in [*Biomedical Optics 2016*], *Biomedical Optics 2016* , OTu2C.7, Optical Society of America (2016).

[6] Aguénounon, E., Dadouche, F., Uhring, W., Ducros, N., and Gioux, S., "Single snapshot imaging of optical properties using a single-pixel camera: a simulation study," *Journal of Biomedical Optics* **24**(7), 1 – 6 (2019).

[7] Rousset, F., Ducros, N., Farina, A., Valentini, G., D'Andrea, C., and Peyrin, F., "Adaptive Basis Scan by Wavelet Prediction for Single-pixel Imaging," *IEEE Transactions on Computational Imaging* (2016).

[8] Mousavi, A., Patel, A. B., and Baraniuk, R. G., "A deep learning approach to structured signal recovery," *CoRR* **abs/1508.04065** (2015).

[9] Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., and Ashok, A., "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (June 2016).

[10] Higham, C., Murray-Smith, R., Padgett, M., and Edgar, M., "Deep learning for real-time single-pixel video," *Scientific Reports.* (Feb 2018).

[11] Hochreiter, S. and Schmidhuber, J., "Long short-term memory," *Neural computation* **9**, 1735–80 (12 1997).

[12] Chung, J., Gülçehre, Ç., Cho, K., and Bengio, Y., "Empirical evaluation of gated recurrent neural networks on sequence modeling," *CoRR* **abs/1412.3555** (2014).

[13] Sankaranarayanan, A. C., Herman, M. A., Turaga, P., and Kelly, K. F., "Enhanced compressive imaging using model-based acquisition: Smarter sampling by incorporating domain knowledge," *IEEE Signal Processing Magazine* **33**, 81–94 (Sept 2016).

[14] Zhang, Z., Wang, X., Zheng, G., and Zhong, J., "Hadamard single-pixel imaging versus fourier single-pixel imaging," *Opt. Express* **25**, 19619–19639 (Aug 2017).

[15] Lorente Mur, A., Ochoa, M., Cohen, J. E., Intes, X., and Ducros, N., "Handling negative patterns for fast single-pixel lifetime imaging," in [*Molecular-Guided Surgery: Molecules, Devices, and Applications V*], Pogue, B. W. and Gioux, S., eds., **10862**, 16 – 25, International Society for Optics and Photonics, SPIE (2019).

[16] Li, C., *An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing*, PhD thesis, Rice University (2010).

[17] Baraniuk, R. G., Goldstein, T., Sankaranarayanan, A. C., Studer, C., Veeraraghavan, A., and Wakin, M. B., "Compressive video sensing: Algorithms, architectures, and applications," *IEEE Signal Processing Magazine* **34**, 52–66 (Jan 2017).

[18] Ducros, N., Lorente Mur, A., and Peyrin, F., "A completion network for compressed acquisition," in [*IEEE 17th International Symposium on Biomedical Imaging (accepted)*], (October 2020).

[19] Lorente Mur, A., Ducros, N., and Peyrin, F., "Recurrent neural networks for compressive video reconstruction," in [*IEEE 17th International Symposium on Biomedical Imaging (accepted)*], (October 2020).

[20] Ballas, N., Yao, L., Pal, C., and Courville, A., "Delving deeper into convolutional networks for learning video representations," *arXiv preprint arXiv:1511.06432* (2015).

[21] Soomro, K., Zamir, A. R., Shah, M., Soomro, K., Zamir, A. R., and Shah, M., "Ucf101: A dataset of 101 human actions classes from videos in the wild," *CoRR* , 2012.

[22] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A., "Automatic differentiation in pytorch," (2017).

[23] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *CoRR* **abs/1412.6980** (2014).

[24] Ducros, N. and Rousset, F., "Single-Pixel Image Reconstruction Toolbox (SPIRiT) Version 2.1." https://github.com/nducros/SPIRIT/ (2020).

[25] Shahzad, A., Walter, N., Malik, A. S., Saad, N. M., and Meriaudeau, F., "Multispectral venous images analysis for optimum illumination selection," in [*2013 IEEE International Conference on Image Processing*], 2383–2387 (Sep. 2013).

[26] Tseng, S.-H., Tzeng, S.-Y., Liaw, Y.-K., Hsu, C.-K., Lee, J., and Chen, W.-R., "Noninvasive evaluation of collagen and hemoglobin contents and scattering property of in vivo keloid scars and normal skin using diffuse reflectance spectroscopy: pilot study," *Journal of Biomedical Optics* **17**(7), 1 – 12 (2012).