

JOINT RECONSTRUCTION OF COMPRESSED MULTI-VIEW IMAGES

Xu Chen

University of Illinois at Chicago
Electrical and Computer Engineering
xchen27@uic.edu

Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4)
pascal.frossard@epfl.ch

ABSTRACT

This paper proposes a distributed representation algorithm for multi-view images that are jointly reconstructed at the decoder. Compressed versions of each image are first obtained independently with random projections. The multiple images are then jointly reconstructed by the decoder, under the assumption that the correlation between images can be represented by local geometric transformations. We build on the compressed sensing framework and formulate the joint reconstruction as a l_2 - l_1 optimization problem. It tends to minimize the MSE distortion of the decoded images, under the constraint that these images have sparse and correlated representations over a structured dictionary of atoms. Simulation results with multi-view images demonstrate that our approach achieves better reconstruction results than independent decoding. Moreover, we show the advantage of structured dictionaries for capturing the geometrical correlation between multi-view images.

Index Terms— compressed sensing, correlation model, stereo images, structured dictionaries, joint reconstruction

1. INTRODUCTION

The advent of distributed architectures has recently raised many interesting research questions for the efficient representation of information in correlated signals captured independently by multiple sensors. In particular, we consider a vision sensor network, in which a number of distributed nodes acquire data and report it to a central collection point which reconstructs the multiple images. In such networks, communication energy and bandwidth are often scarce resources, which imposes important constraints in terms of communication costs or bit rate. Distributed source coding [1] permits to achieve substantial rate savings when the correlation between the source can be modeled accurately, which is not trivial in images.

In this paper, we adopt a different approach where sensors perform random measurements on the multi-view images, and the joint decoder tries to estimate the underlying correlation between images for efficient reconstruction. We build on the compressed sensing framework, and we formulate the joint reconstruction problem as a l_2 - l_1 optimization problem. The

decoder tries to minimize the distortion of the reconstructed images. At the same time, it imposes that the images have sparse representations in a structured dictionary of geometric atoms, and that these representations are correlated by local geometrical transformations such as translations or rotation.

Compressed sensing has been extended recently to distributed scenarios, but most recent works consider that the multiple observations are noisy representations of the same underlying signal, or that signals have the same sparse support [2]. Our work rather addresses the problem where the observations represent slightly transformed versions of a common signal, as well as innovation. Research work in super-resolution [3] or recently compressive coding aperture problems [4] similarly try to exploit multiple correlated signals for effective joint reconstruction. They however do not exploit the interesting properties of redundant structured dictionaries for joint reconstruction of signals that are related by local geometrical transformations.

The rest of the paper is organized as follows: Compressed sensing and structured dictionaries are discussed in Section 2. We propose our novel joint reconstruction approach for multi-view images in Section 3. In Section 4, we demonstrate the superiority of our algorithm over traditional approaches for the joint reconstruction of multi-view images.

2. PRELIMINARIES

2.1. Compressed sensing

The compressed sensing framework [5] provides an interesting new paradigm for the sampling and reconstruction of signals, where a small number of random measurements are sufficient for representing signals that have a sparse support in some basis ψ . In particular, let us consider a very high dimensional signal $X \in R^N$. The recovery of the signal can be achieved with high accuracy from a much smaller dimensional observation b in a measurement basis ϕ by the minimization of the l_1 norm $\|\alpha\|_1$ of the coefficient vector α in the sparse basis ψ , under the constraint $\|b - \phi\psi\alpha\|_2^2 \leq \varepsilon$. The observation matrix ϕ is of the size $M \times N$ ($M \ll N$), and b is $M \times 1$ vector where $M \ll N$. The measurement matrix ϕ can be random matrix, built on random Gaussian or binary

vectors. Proper reconstruction can be obtained when ϕ satisfies the Restricted Isometry Property (RIP) [6], or when ψ and ϕ are sufficiently incoherent. Finally, the basis ψ is usually a subset of an overcomplete dictionary, which provides a K -sparse representation of X . Namely X can be written as a linear combination of K atoms (i.e., basis functions).

2.2. Sparse approximations with structured dictionaries

Given a redundant dictionary of atoms $\mathcal{D} = \{\psi_k\}, k = 1 \dots P$, in the Hilbert space \mathcal{H} , we say that an image I_1 has a sparse representation in \mathcal{D} if it can be approximated by a linear combination of a small number of vectors from \mathcal{D} that are represented in the matrix ψ_1 . Therefore, sparse approximation of the image I_1 can be expressed as $I_1 = \psi_1\alpha_1 + \eta_1$, where η_1 is the approximation error.

In particular, one could use a redundant structured dictionary \mathcal{D} built by applying geometrical transformations to a generating function, for example. A structured and possibly redundant dictionary $\mathcal{D} = \{\psi_\gamma, \gamma \in \Gamma\}$ spanning the input space. The atoms in a *structured* dictionary are constructed by applying geometric transformations to a generating mother function denoted by ψ . A geometric transformation $\gamma \in \Gamma$ can be represented by a unitary operator $U(\gamma)$, so that a structured dictionary takes the following form, $\mathcal{D} = \{\psi_\gamma U(\gamma)\psi, \gamma \in \Gamma\}$. With digital images, the simplest transformation γ_i , applied to the i -th atom, may be one of the following three types.

- *Translation* by $t_i = [t_{ix} \ t_{iy}]^\top$. $U(t_i)$ moves the generating function across the image i.e., $U(t_i)\psi(x, y) = \psi(x - t_{ix}, y - t_{iy})$.
- *Rotation* by θ_i . $U(\theta_i)$ rotates the generating function by angle θ_i i.e., $U(\theta_i)\psi(x, y) = \psi(\cos(\theta_i)x + \sin(\theta_i)y, \cos(\theta_i)y - \sin(\theta_i)x)$.
- *Anisotropic scaling* by $a_i = [a_{ix} \ a_{iy}]^\top$. $U(a_i)$ scales the generating function anisotropically in the two directions i.e., $U(a_i)\psi(x, y) = \psi(\frac{x}{a_{ix}}, \frac{y}{a_{iy}})$.

Composing all the above transformations yields a transformation with parameters $\gamma_i = \{t_i, a_i, \theta_i\} \in \Gamma$, which denotes a synthesis of translations, anisotropic scalings and rotations. That forms a matrix Ψ , where each row correspond to an atom in \mathcal{D} . As it can be seen from above discussion, structured dictionaries provide an easy way for the joint representation of images that are correlated by geometrical transforms since the indices of atoms in the dictionary contains the information about the geometry of the basis functions. Basically, if I_1 and I_2 are correlated under some local geometrical transforms, one could write $I_2 \approx F(\psi_1)\alpha_{21} + \psi_2\alpha_{22}$, where ψ_1 represents atoms that are used for the sparse approximation of I_1 and F represents the geometrical transform that results in other atoms in the same dictionary \mathcal{D} .

3. JOINT RECONSTRUCTION

We consider now the case where compressed representation of multi-view images have been generated in a distributed way. The decoder jointly reconstructs the multi-view images in exploiting the correlation between the different signals. In particular, we assume that this correlation takes the form of local geometrical transformations. In particular, we assume that the correlation between images could be captured by geometrical transformation of atoms: $I_2 \approx F(\psi_1)\alpha_{21} + \psi_2\alpha_{22}$.

Each sensor performs independent measurements for each image $b_i = \phi_i I_i$, and the decoder jointly reconstructs the multiple image using the correlation model above, when a structured dictionary Φ is used at decoder. Without loss of generality, we assume that we have three correlated images, I_1, I_2 and I_3 . Under the correlation model, we can write:

$$I_1 \approx \psi_1\alpha_1, \quad (1)$$

$$I_2 \approx F_{12}(\psi_1)\alpha_{21} + \psi_2\alpha_{22}, \quad (2)$$

$$I_3 \approx F_{13}(\psi_1)\alpha_{31} + F_{23}(\psi_2)\alpha_{32} + \psi_3\alpha_{33} \quad (3)$$

We can cast the reconstruction problem as a $l^2 - l^1$ optimization problem where we jointly minimize the reconstruction error and the number of components used in the reconstruction:

$$(F^*, \alpha^*) = \arg \min_{F_i, \alpha_i} \|B - \phi A \alpha\|_2^2 + \tau \|\alpha\|_1, \quad (4)$$

where $B = [b_1, b_2, b_3]^T$,

$$\alpha = [\alpha_1, \alpha_{21}, \alpha_{22}, \alpha_{31}, \alpha_{32}, \alpha_{33}]^T,$$

$$A = \begin{pmatrix} \psi_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & F_{12}(\psi_1) & \psi_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & F_{13}(\psi_1) & F_{23}(\psi_2) & \psi_3 \end{pmatrix},$$

and ψ_1, ψ_2, ψ_3 are the subset dictionaries for the stereo images I_1, I_2 and I_3 .

The problem is equivalent to minimize the norm of the coefficient vector,

$$\min_{F_i, \psi_i, \alpha} \|\alpha\|_1, \quad (5)$$

subject to $\frac{1}{2} \|B - \phi A \alpha\|_2^2 \leq \varepsilon$.

Such an optimization problem is similar to a basis pursuit formulation and can be solved by traditional linear programming techniques [7].

Finally, note that the above optimization problem could be formulated in a similar manner when images I_2 or I_3 are assumed to be the reference images, i.e., they have not been transformed. For example, when the problem is centered on I_2 , we have

$$I_1 \approx \psi_1\alpha_{11} + F_{12}(\psi_2)\alpha_{12}, \quad (6)$$

$$I_2 \approx \psi_2\alpha_2, \quad (7)$$

$$I_3 \approx F_{23}(\psi_2)\alpha_{32} + \psi_3\alpha_{33} \quad (8)$$

4. EXPERIMENTAL RESULTS

4.1. Setup

For the joint reconstruction, we choose to use the structured dictionary proposed in [8]. It is built by applying geometric transformations to a generating mother function g , the atoms are, therefore, indexed by a string γ composed of five parameters: translation \vec{t} , anisotropic scaling \vec{a} and rotation θ . Any atom in our dictionary can be finally be expressed in the following form:

$$\hat{g} = \frac{2}{\sqrt{3\pi}}(4\hat{g}_1^2 - 2)\exp(-(\hat{g}_1^2 + \hat{g}_2^2)),$$

$$\hat{g}_1 = \frac{\cos(\theta)(x - t_1) + \sin(\theta)(y - t_2)}{a_1},$$

$$\hat{g}_2 = \frac{\cos(\theta)(y - t_2) + \sin(\theta)(x - t_1)}{a_2}$$

We build the overcomplete dictionary ψ as follows: Let x, y vary from 1 to 32 respectively. Let θ vary from 0 to π with the step size $\pi/18$, $a_1 = a_2 = 4$. Therefore, the full dictionary contains totally 19456 atoms. Specifically, the indices from $1024 \times (k-1) + 1$ to $1024 \times k$ corresponds to the atoms with $\theta = (k-1)\pi/18$. We select the sub-dictionary ψ_1 by computing the absolute value of inner products of the measurement b for image I_1 and $\phi\psi$ and select the first W atoms with the maximum results.

In our experiments, we consider translation and rotation in the correlation between images. We then solve the optimization problems above using SPGL1 toolbox at [9], which is a Matlab solver for large-scale one-norm regularized least squares relying on basis pursuit. For all the simulations, the tolerance $\epsilon = 0.01$ and the iteration times is set to be 500. The stereo images are available at the database [10], which are reshaped into the size of 64 by 64 with bilinear interpolation and jointly reconstructed with 32 by 32 subimages.

4.2. Joint reconstruction performance

We first compare the performance of the joint reconstruction algorithm, where the correlation is modeled as local or global transformations between images. For local transforms, each atom can undergo different transformation. In the latter, we consider that all the atoms in ψ_i undergo the same transformation, which is chosen as the transformation that appears the most frequently among pairs of atoms in different images. We compare the performance to independent reconstruction, where each image is reconstructed independently with classical compressive sensing. It uses the same number of measurements as the joint reconstruction algorithm, and the same number of atoms for the reconstruction. The illumination of each pixel value has been normalized into the range from 0 to 1.

Figures 1 and 2 shows the average MSE distortion for reconstruction based on the optimization problems of Eqs. (1-3)

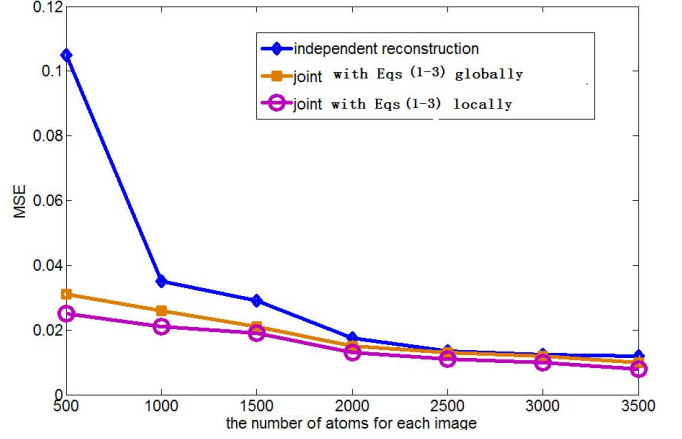


Fig. 1. MSE distortion versus different number of atoms, with the optimization problem of Eqs. (1-3).

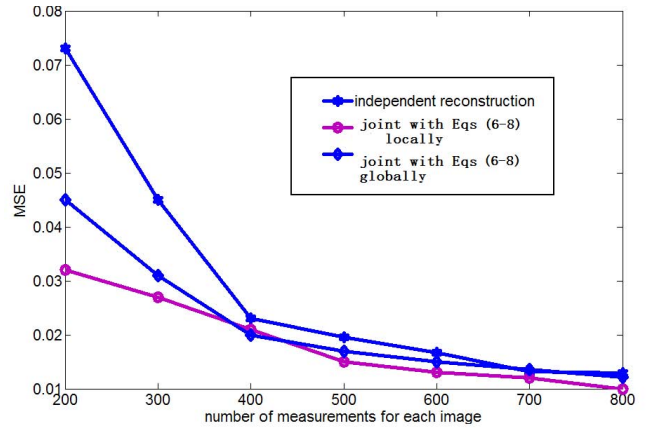


Fig. 2. MSE distortion versus different number of measurements, with the optimization problem of Eqs. (6-8).

and (6-8) respectively, as a function of the number of measurements per image. It can be seen that our joint reconstruction approach is much superior to independent reconstruction especially when a small number of atoms or measurements per image is utilized due to a sparser representation when the geometrical correlation is considered. Moreover, local transformations of atoms yield lower MSE than global transformations. It has to be noted however that global transformations require less computation time.

We then illustrate in Fig. 3 the proposed joint reconstruction scheme when the first or the second image is used as the reference image. This respectively corresponds to the optimization problems given by Eqs (1-3) and (6-8) above. We select 1800 atoms and 600 measurements for each image. It can be seen that the joint reconstruction improves the quality of all the images, except the reference image. It can be seen that choosing image 1 as reference image, the reconstruction of image 3 is the best of the three images due to the corre-

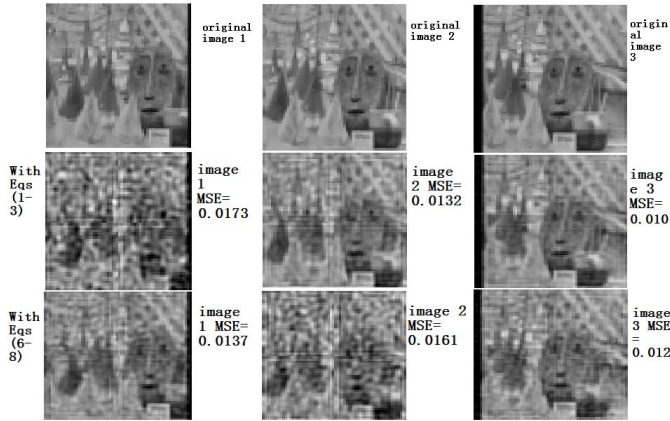


Fig. 3. The comparison of reconstruction results and MSE with optimization problems of Eqs. (1-3) and (6-8).

lation models with image 1 and image 2. The proposed approach can be easily extended to jointly reconstruction of L images and significant improvements are achieved for $N - 1$ images.

4.3. Influence of reconstruction basis

Wavelet bases are widely used for reconstruction of images in the compressed sensing framework [5]. Comparing with a wavelet basis, our structured dictionary has two main superiorities:

1. Our structured dictionary contains translation and rotation, which provides great convenience and saves significant computation time when multi-view images are jointly reconstructed although we did not show here due to the limitation of the space.
2. Given the same limited number of atoms and measurements, the atoms from the structured dictionary generally provides higher approximation performance for images at low rate

We illustrate in Table I the performance of independent reconstruction with a haar wavelet basis and joint reconstruction with our structured dictionary. We consider three multiview images with 1024 atoms for the reference image.

Table 1. Comparison of mean square error for reconstruction with Haar wavelet basis, and our dictionary

atoms	haar wavelet	gaussian (Eqs. (1-3))
400 measurements	0.046	0.032
500 measurements	0.039	0.028
600 measurements	0.025	0.018
700 measurements	0.021	0.013
800 measurements	0.019	0.009

5. CONCLUSION

In this paper, we propose a novel approach for joint reconstruction of multiview images that have been compressed independently. We build on the compressed sensing framework, and introduce a geometrical correlation model for the joint reconstruction of the images at the decoder. In particular, we assume that the images have a sparse decomposition over a structured dictionary of geometrical atoms. Furthermore, we assume the decompositions of different images are related by local geometrical transformations of atoms, within the same dictionary. We formulate the joint reconstruction as a l_2 - l_1 optimization problem, and experimental results demonstrate the superiority of our approach over independent reconstruction of multi-view images. Future work will consider the computational complexity of the proposed approach.

6. REFERENCES

- [1] D.Slepian and J.K.Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, 1973.
- [2] M.F.Duarte, S.Sarvotham, D.Baron, M.B.Wakin, and R.G.Baraniuk, "Distributed compressed sensing of jointly sparse signals," *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on AB - ER -*, pp. 1537–1541, 2005.
- [3] R.C.Hardie, K.J.Barnard, and E.E.Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, pp. 1621–1633, 1997.
- [4] R.F.Marcia and R.M.Willett, "Compressive coded aperture video reconstruction," *16th European Signal Processing Conference*, 2008.
- [5] D.L.Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [6] J.Haupt and R.Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. on Information Theory*, vol. 52, pp. 4036–4048, 2006.
- [7] D.Baron, M.F.Duarte, S.Sarvotham, M.B.Wakin, and R.G.Baraniuk, "An information theoretic approach to distributed compressed sensing," *Proceedings of the 43d Allerton Conference on Communication, Control, and Computing*, 2005.
- [8] R.M.Figueras, P.Vanderghyest, and P.Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. Image Processing*, vol. 15, pp. 726–739, 2006.
- [9] "http://www.cs.ubc.ca/labs/scl/spgl1/," *Scientific Computing Laboratory in University of British Columbia*, 2008.
- [10] "http://cat.middlebury.edu/stereo/data.html," *Middlebury Stereo Vision Page*.