ARTICLE IN PRESS

J. Vis. Commun. Image R. xxx (2012) xxx–xxx

Correlation estimation from compressed images ☆

Vijayaraghavan Thirumalai *, Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory (LTS4), Lausanne 1015, Switzerland

A R T I C L E   I N F O

A B S T R A C T

This paper addresses the problem of correlation estimation in sets of compressed images. We consider a framework where the images are represented under the form of linear measurements due to low complexity sensing or security requirements. We assume that the images are correlated through the displacement of visual objects due to motion or viewpoint change and the correlation is effectively represented by optical flow or motion field models. The correlation is estimated in the compressed domain by jointly processing the linear measurements. We first show that the correlated images can be efficiently related using a linear operator. Using this linear relationship we then describe the dependencies between images in the compressed domain. We further cast a regularized optimization problem where the correlation is estimated in order to satisfy both data consistency and motion smoothness objectives with a Graph Cut algorithm. We analyze in detail the correlation estimation performance and quantify the penalty due to image compression. Extensive experiments in stereo and video imaging applications show that our novel solution stays competitive with methods that implement complex image reconstruction steps prior to correlation estimation. We finally use the estimated correlation in a novel joint image reconstruction scheme that is based on an optimization problem with sparsity priors on the reconstructed images. Additional experiments show that our correlation estimation algorithm leads to an effective reconstruction of pairs of images in distributed image coding schemes that outperform independent reconstruction algorithms by 2–4 dB.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

In recent years, the increasing popularity of vision sensor networks has led to the generation of huge volume of visual information. This creates the need for effective information processing systems that are able to efficiently compress, analyze and store highly redundant information streams captured by multiple devices. Distributed processing solutions become highly attractive in such a context, as they permit to reduce the communication and computational power requirements in the sensors. The visual information is typically compressed and transmitted independently from each sensor node to a common decoder that jointly processes the correlated information streams. The inter-sensor communication needs are relaxed and the computational burden is shifted to the decoder. The estimation of the image correlation at decoder becomes however crucial in such distributed settings for image reconstruction or analysis tasks.

In this paper, we consider the problem of correlation estimation in a framework where multiple sensors transmit compressed images that have been obtained by a small number of linear projections of the original images, as illustrated in Fig. 1. Such linear projections typically represent simple measurements in low complexity sensing systems [2,3]. We propose a novel solution for correlation estimation at the joint decoder where the analysis is performed directly in the compressed domain in order to avoid expensive image reconstruction tasks. This is especially useful for the analysis applications that do not target image reconstruction. We assume that the correlation between images corresponds to camera or object motion; this can be efficiently represented by optical flow or motion field models. We show that such a correlation model can be described by a linear operator and we further analyze in detail the effect of such operator in the compressed domain. Later, we cast the correlation estimation as a regularized energy minimization problem with constraints on data consistency as well as consistency of the motion field. In particular, we regularize the correlation model such that the motion values in neighboring pixels are similar except at image discontinuities. Such an optimization problem can be solved by Graph Cuts algorithms.

We analyze in details the performance of our novel correlation estimation framework. In particular, we study the penalty in the correlation estimation that is due to working in the compressed

---

* Corresponding author.
E-mail addresses: vijayaraghavan.thirumalai@epfl.ch (V. Thirumalai), pascal. frossard@epfl.ch (P. Frossard).

ARTICLE IN PRESS

2   *V. Thirumalai, P. Frossard / J. Vis. Commun. Image R. xxx (2012) xxx–xxx*

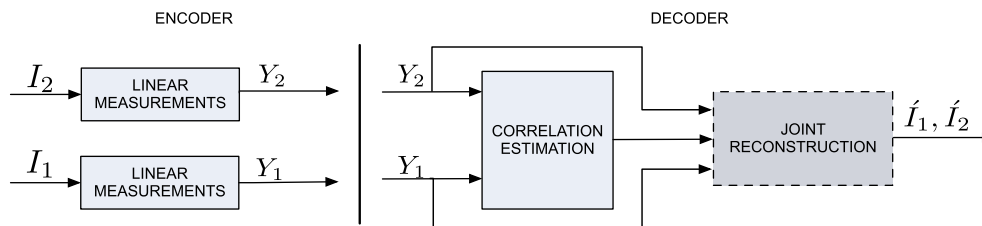ENCODER   DECODER

Fig. 1. Schematic representation of the proposed scheme. The images I_1 and I_2 are correlated through displacement of scene objects due to viewpoint change or motion of scene objects. The correlation model is estimated directly in the compressed domain without any intermediate image reconstruction. The correlation information can be used for an optional joint reconstruction.

domain as opposed to the original image domain as in traditional correlation estimation problems. We show that the penalty decreases when the number of measurements increases and that our algorithm tends to the optimal correlation estimate at high measurement rate. Extensive simulations in distributed stereo and video imaging applications confirm that the proposed solution provides effective estimates of the relative motion between images and even competes with solutions that implement expensive image reconstruction prior to correlation estimation. We finally study the performance of a novel joint reconstruction algorithm that uses our correlation estimates for decoding pairs of images. The joint reconstruction is cast as an optimization problem where the reconstructed images have to satisfy sparsity priors as well as consistency with both the measurements and the correlation estimates. We solve this joint reconstruction problem by effective proximal splitting methods and show that accurate correlation estimation in distributed image representation permits to outperform independent decoding solutions in terms of image quality.

The rest of the paper is organized as follows. Section 2 briefly overviews the related work. In Section 3, we describe the proposed framework and show how the correlation estimation problem carries out to the compressed domain. Section 4 describes the proposed correlation estimation algorithm and its performance are analyzed in details on Section 5. In Section 6 we draw some concluding remarks.

2. Related work

This section describes the literature related to the framework proposed in this paper. We first present sensing solutions based on linear measurements. We then review the correlation estimation algorithms in distributed image representation systems. Finally, we discuss the most relevant works about joint reconstruction of correlated images.

In recent years, signal acquisition based on random projections received a significant attention in many applications like medical imaging, compressive imaging and even sensor networks. Donoho [2] and Candes et al. [3] show that the small number of linear measurements contain enough information to reconstruct a sparse or a compressible signal. In particular, they show that if a signal has a sparse representation in one basis then it can be recovered from a small number of linear measurements taken on another (random) basis that is incoherent with the first one. Essentially, if the signal is K-sparse (i.e., if the signal contains K significant components), then one need approximately cK linear measurements (typically $c = 3$ or 4) to reconstruct the signal with high probability [4]. Such results open the door to novel low complexity sensing solutions where the computational complexity for signal reconstruction or analysis is pushed to the decoder. These ideas have been applied to image acquisition [5–7] and later extended to video sequences [8–11]. The effect of measurement quantization and the lossy compression of linear measurements has been studied in [12].

One of the key characteristics in imaging applications resides in the high correlation between multi-view images or successive images in a video sequence. The correlation could be exploited for effective reconstruction of image sets or for joint analysis tasks in distributed systems. Duarte et al. [13,14] have proposed different correlation models for the distributed compression of correlated signals from linear measurements. In particular, they introduce three joint sparsity models (JSM) in order to exploit the inter-signal correlation for the joint reconstruction. They are respectively described by (i) JSM-1, where the signals share a common sparse support plus a sparse innovation part specific to each signal, (ii) JSM-2, where the signals share a common sparse support with different coefficients, and (iii) JSM-3 with a non-sparse common signal with an individual sparse innovation in each signal. These correlation models permit an effective joint reconstruction with a small number of measurements compared to the independent reconstruction. These simple joint sparsity models are however not ideal for multi-view images or video sequences, as the correlation model in such scenarios is usually given in the form of disparity or motion vectors respectively. The authors in [15–17] have proposed a distributed joint reconstruction scheme for video sequences based on linear measurements. These schemes split the video sequences into key frames and compressed sensing (CS) frames. The random measurements are computed independently for each compressed sensing frame and are transmitted to the joint decoder. The key frames are intra coded and the joint decoder builds the side information from the intra coded key frames. The generated side information is then used to decode the CS frame by solving an optimization problem which assumes that the prediction error for the CS frame is sparse in an orthonormal basis [16] or block-based adaptive dictionary [17]. Similar ideas have been used by Trocan et al. [18,19] for distributed multi-view compression where all the images are however given in the form of linear measurements. The joint decoder first reconstructs all the views by solving a regularized optimization problem. Then, the independently reconstructed views are used to estimate the underlying correlation model in the form of disparity image. The disparity image is used to jointly reconstruct all the views in a multistage refinement framework in which each refinement stage reconstructs a view by assuming that the prediction error is sparse in a dual tree wavelet basis. In our previous work [20], we have proposed a methodology to estimate the correlation between pairs of frames where one image serves as a reference image. Unfortunately, reconstructing the reference images in a compressed measurements framework typically requires methods based on solving l_2–l_1 optimization problems that are highly complex. The works in the literature typically use reference frames that are encoded as intra-frames or reconstructed with complex optimization tools prior to correlation estimation. In this paper, we rather propose to avoid the explicit reconstruction of the images and directly estimate the correlation in the compressed domain. In general, it permits to reduce the computational complexity at the decoder, especially in applications where image reconstruction is not necessary.

ARTICLE IN PRESS

V. Thirumalai, P. Frossard / J. Vis. Commun. Image R. xxx (2012) xxx–xxx 3

Finally, other works have recently addressed the problem of joint reconstruction of correlated images given in compressed form. For example, Park et al. [21] have proposed an image registration and joint reconstruction algorithm for multi-view images based on manifold lifting. The underlying correlation between views is exploited by constructing an image appearance manifold in which the images represent sample points on the manifold; these points are controlled by a few camera parameters (e.g., rotation, translation etc.). By knowing the initial camera positions the images are jointly reconstructed based on an l_2–l_1 optimization framework. Then, the reconstructed scene is used to refine the camera parameters and thus both the camera positions and the scene are jointly estimated using alternating minimization techniques. In another framework [22], the authors have proposed a joint reconstruction scheme based on a regularized optimization framework. The two regularization terms encourage sparse priors of multi-view images and the difference between images. However, the correlation between images is not efficiently exploited as the correlation model in multi-view image settings is usually given in the form of disparity image and not as a sparsity prior of the signal differences. The joint reconstruction scheme proposed in this paper rather builds the correlation model in the form of disparity or motion field and thus facilitates an efficient joint image representation. Furthermore, the correlation model is built directly from the linear measurements in the compressed domain and thus avoids the computational complexity of reconstructing the reference images. The proposed scheme is based on low complexity linear measurements and it provides an interesting flexible solution for distributed processing in vision sensors, targeting applications like object detection, distributed rendering or distributed joint signal reconstruction.

3. Distributed representation of correlated images

3.1. Framework

We consider a framework where the images represent a scene at different time instants or from different viewpoints. For the sake of clarity, we consider a pair of images I_1 and I_2 (with resolution $N = N_1 \times N_2$) but the framework extends to larger number of images. These images are represented by linear measurements that correspond to the projection of the image pixel values on a set of coding vectors. Typically, the coding vectors can be constructed from Gaussian or Bernoulli distributions [2] or with a block structure [7,23] for easier handling and fast sampling of large images. The measurements are transmitted to a joint decoder that can estimate the correlation between the compressed images and possibly perform a joint reconstruction of the image set. The framework is illustrated in Fig. 1.

In more details, the sensors process images row by row. Let $I_{1,k}$ and $I_{2,k}$ represent the kth row of the images I_1 and I_2 respectively, and $Y_{1,k}$ and $Y_{2,k}$ represent the linear measurements computed from $I_{1,k}$ and $I_{2,k}$ using the measurement matrices ϕ_1^k and ϕ_2^k respectively. The measurements $Y_{1,k}$ and $Y_{2,k}$ are computed as

$$
\begin{aligned}
Y_{1,k} &= \phi_1^k I_{1,k}^T, \quad \forall k = 1, 2, \ldots, N_1, \\
Y_{2,k} &= \phi_2^k I_{2,k}^T, \quad \forall k = 1, 2, \ldots, N_1,
\end{aligned}
\tag{1}
$$

where $(\cdot)^T$ denotes the transpose operator. It should be noted that ϕ_1^k and ϕ_2^k are of dimensions $M \times N_2$, where $M \ll N_2$ is the number of measurements computed for each row in the image. From Eq. (1) it is easy to check that the measurements $Y_1 = [Y_{1,1}, Y_{1,2}, \ldots, Y_{1,N_1}]^T$ and $Y_2 = [Y_{2,1}, Y_{2,2}, \ldots, Y_{2,N_1}]^T$ can be computed as

$$
\begin{bmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ Y_{i,N_1} \end{bmatrix} = \Phi_i \underbrace{\begin{bmatrix} I_{i,1}^T \\ I_{i,2}^T \\ \vdots \\ I_{i,N_1}^T \end{bmatrix}}_{I_i}, \quad \forall i \in \{1, 2\},
\tag{2}
$$

where Φ_i is the measurement matrix used to sample the ith image. It is represented as

$$
\Phi_i = \begin{bmatrix} \phi_i^1 & 0 & \ldots & 0 \\ 0 & \phi_i^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \phi_i^{N_1} \end{bmatrix}_{K \times N}, \quad \forall i \in \{1, 2\},
\tag{3}
$$

where $K = MN_1$, $N = N_1 N_2$ and K/N represents the measurement rate.

3.2. Correlation model

In the above settings, the correlation between images is mainly explained by the relative displacement of objects in the scene. This can be effectively modeled by the optical flow that determines the amount of displacement of objects or pixels in different images. We show now, how such a correlation model can be described by a linear operator. Let $\mathbf{m}^h$ and $\mathbf{m}^v$ represent the horizontal and vertical motion components. As the visual objects in the images I_1 and I_2 are displaced, it is likely that the pixel at position $\mathbf{z} = (k, l)$ in one image moves to $\mathbf{z}' = (k + \mathbf{m}^h(k,l), l + \mathbf{m}^v(k,l))$ in the second image. Thus, the images I_1 and I_2 can be simply related by a linear operator $\mathcal{T}$ that changes the coordinate system from (k, l) in the first image to $(k + \mathbf{m}^h(k,l), l + \mathbf{m}^v(k,l))$ in the second image, i.e.,

$$
\begin{aligned}
I_2 &= \mathcal{T}\{I_1\}, \\
I_{2,k}(l) &= I_{1,(k+\mathbf{m}^h(k,l))}(l + \mathbf{m}^v(k,l)).
\end{aligned}
\tag{4}
$$

For mathematical convenience we use an equivalent representation of Eq. (4) in the form of matrix multiplication:

$$
I_{2,k}^T = A^k \underbrace{\begin{bmatrix} I_{1,1}^T \\ I_{1,2}^T \\ \vdots \\ I_{1,N_1}^T \end{bmatrix}}_{I_1}, \quad \forall k = 1, 2, \ldots, N_1,
\tag{5}
$$

where A^k is a matrix of dimensions $N_2 \times N_1 N_2$ whose entries are determined by the horizontal and vertical components of the motion field in the kth row of pixels, i.e., $\mathbf{m}^h(k,\cdot)$ and $\mathbf{m}^v(k,\cdot)$. The elements of the matrix A^k are given by

$$
A^k(l, l + \beta_1 + \beta_2 N_2) = \begin{cases} 1 & \text{if } \mathbf{m}^h(k,l) = \beta_1, \ \mathbf{m}^v(k,l) = \beta_2 \\ 0 & \text{otherwise}. \end{cases}
\tag{6}
$$

If $l + \beta_1 + \beta_2 N_2 > N_1 N_2$ (e.g., at image boundaries), we set $l + \beta_1 + \beta_2 N_2 = N_1 N_2$ so that the dimensions of the matrix A^k stays $N_2 \times N_1 N_2$. It is easy to check that the matrix A^k formed using Eq. (6) contains only one '1' in each row; this implies $I_{2,k}(l) = I_{1,k+\beta_1}(l + \beta_2)$ if $A^k(l, l + \beta_1 + \beta_2 N_2) = 1$. The action of the matrix A^k shifts the pixels in I_1 according to the motion given as $\mathbf{m}^h(k,\cdot)$ and $\mathbf{m}^v(k,\cdot)$ and forms an estimate of the image $I_{2,k}$. It should be noted that the matrix A^k is completely determined by the kth row of motion vectors $\mathbf{m}^h(k,\cdot)$ and $\mathbf{m}^v(k,\cdot)$.

The relation given in Eq. (5) can be extended to all rows of the image I_2. The images I_1 and I_2 are finally related by a linear operator A such that $I_2 = AI_1$ which can be written as

ARTICLE IN PRESS

4   V. Thirumalai, P. Frossard / J. Vis. Commun. Image R. xxx (2012) xxx–xxx

$$\underbrace{\begin{bmatrix} I_{2,1}^T \\ I_{2,2}^T \\ \vdots \\ I_{2,N_1}^T \end{bmatrix}}_{I_2} = \underbrace{\begin{bmatrix} A^1 \\ A^2 \\ \vdots \\ A^{N_1} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} I_{1,1}^T \\ I_{1,2}^T \\ \vdots \\ I_{1,N_1}^T \end{bmatrix}}_{I_1}. \qquad (7)$$

This relation is illustrated on the lefthand side of Fig. 2.

3.3. Correlation in the measurement domain

We now extend the above correlation model in the compressed domain. Without loss of generality, we first assume that the measurements Y_1 and Y_2 can be related by a linear transformation B, i.e.,

$$\underbrace{\begin{bmatrix} Y_{2,1} \\ Y_{2,2} \\ \vdots \\ Y_{2,N_1} \end{bmatrix}}_{Y_1} = \underbrace{\begin{bmatrix} B^1 \\ B^2 \\ \vdots \\ B^{N_1} \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} Y_{1,1} \\ Y_{1,2} \\ \vdots \\ Y_{1,N_1} \end{bmatrix}}_{Y_2}, \qquad (8)$$

where $B^k, \forall k = 1, 2, \ldots, N_1$ is a matrix with dimensions $M \times MN_1$, i.e., the measurements $Y_{2,k}$ can be related to Y_1 as

$$Y_{2,k} = B^k Y_1, \quad \forall k = 1, 2, \ldots, N_1. \qquad (9)$$

Any two vectors $Y_1, Y_2 \in \mathbb{R}^{MN_1}$ can be related by a linear transformation B as long as $Y_1 \neq \mathbf{0}$, which is the case in our framework. However, this linear transformation B a priori does not have any special form. We are interested in understanding the relation between this matrix and the matrix A that shifts the pixels between images I_1 and I_2. Pre-multiplying Eq. (7) by Φ_2 on both sides, one can write

$$Y_2 = \Phi_2 I_2 = \Phi_2 A I_1. \qquad (10)$$

In addition, by replacing $Y_1 = \Phi_1 I_1$ in Eq. (8), one obtains

$$Y_2 = B Y_1 = B \Phi_1 I_1. \qquad (11)$$

From Eqs. (10) and (11) the relation between B and A can finally be given as

$$B \Phi_1 = \Phi_2 A. \qquad (12)$$

This forms an over-determined system of linear equations, as the number of unknown in matrix B is smaller than the number of equations in $\Phi_2 A$. In this case, the optimal matrix $\widehat{B}$ that minimizes $\|B\Phi_1 - \Phi_2 A\|_2$ is given by

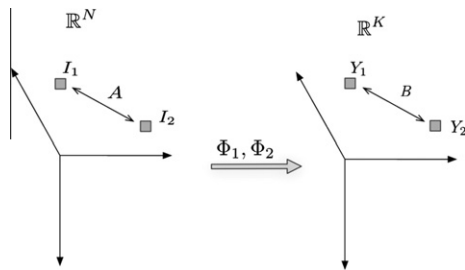$$\widehat{B} = \Phi_2 A \Phi_1^\dagger, \qquad (13)$$

Fig. 2. Illustration of the relation between the matrices A and B. On the left the images $I_1, I_2 \in \mathbb{R}^N$ are related using the matrix A where $N = N_1 N_2$. In the compressed domain, the measurement vectors $Y_1, Y_2 \in \mathbb{R}^K$ are related using the matrix B where $K = MN_1$. The matrices A and B can be related by $B \approx \Phi_2 A \Phi_1^\dagger$ where Φ_i's are the sensing matrices.

where † denotes the pseudo-inverse operator. As the rows in Φ_1 are generally orthonormal, the pseudo-inverse can be computed using the transpose operator, i.e., $\Phi_1^\dagger = \Phi_1^T$. Substituting $\widehat{B} = \Phi_2 A \Phi_1^T$ in Eq. (8) the relation between the measurements becomes

$$Y_2 \approx \Phi_2 A \Phi_1^\dagger Y_1 = \Phi_2 A \Phi_1^T Y_1, \qquad (14)$$

$$Y_{2,k} \approx \phi_2^k A^k \Phi_1^T Y_1, \quad \forall k = 1, 2, \ldots, N_1. \qquad (15)$$

Eq. (15) comes from the fact that measurements are computed across rows of pixels in our framework. The relationship between the matrices A and B is illustrated in Fig. 2, where the matrices A and B are used to relate the points in the original and compressed domains, respectively. In the next section, we propose an algorithm for estimating the correlation model directly from the linear measurements Y_1 and Y_2.

4. Correlation estimation from linear measurements

4.1. Regularized energy minimization problem

We propose in this section a method for estimating the correlation between images from the compressed measurements without any explicit image reconstruction steps. The objective is to compute a flow or motion field that represents the motion between images I_1 and I_2. We denote this flow field as $\mathcal{M} = (\mathbf{m}^h, \mathbf{m}^v)$, where $\mathbf{m}^h$ and $\mathbf{m}^v$ are horizontal and vertical components of the motion, respectively. The problem then consists in finding the value of the flow field $\mathcal{M}$ at each pixel position $\mathbf{z} = (k, l)$ such that the estimated correlation is consistent with the measurement vectors Y_1 and Y_2. At the same time, the motion field has to be piecewise smooth in order to model consistent motion of visual objects. We propose to cast the correlation estimation as a regularized energy minimization problem, where the energy $E(\mathcal{M})$ is composed of a data term $E_d(\mathcal{M})$ and a smoothness term $E_s(\mathcal{M})$. The optimal mapping $\mathcal{M}^*$ is obtained by minimizing the energy function $E(\mathcal{M})$ as

$$\mathcal{M}^* = \arg \min_{\mathcal{M}} E(\mathcal{M}) = \arg \min_{\mathcal{M}} [E_d(\mathcal{M}) + \lambda E_s(\mathcal{M})], \qquad (16)$$

where λ balances the importance of the data and smoothness terms.

We now discuss in more details the components of the energy function. The smoothness term measures the penalty of assigning different motion values to the adjacent pixels. We write it as

$$E_s(\mathcal{M}) = \sum_{\mathbf{z},\mathbf{z}' \in \mathcal{N}} V_{\mathcal{M}}(\mathbf{z}, \mathbf{z}'), \qquad (17)$$

where $\mathbf{z}, \mathbf{z}'$ are neighbour pixels in the 4-pixel neighbourhood denoted by $\mathcal{N}$. The term $V_{\mathcal{M}}(\mathbf{z}, \mathbf{z}')$ is given as

$$V_{\mathcal{M}}(\mathbf{z}, \mathbf{z}') = min(|\mathbf{m}^h(\mathbf{z}) - \mathbf{m}^h(\mathbf{z}')| + |\mathbf{m}^v(\mathbf{z}) - \mathbf{m}^v(\mathbf{z}')|, \tau), \qquad (18)$$

where τ sets an upper level on the smoothness penalty that helps preserving the discontinuities [24].

Next, the data function measures the consistency of a particular motion value for pixel $\mathbf{z}$ with the vectors Y_1 and Y_2. Classically, the accuracy of the motion values is evaluated with the original images [25] and the data cost is typically given as

$$\widetilde{E}_d(\mathcal{M}) = \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} \Delta_{\mathcal{M}}(k, l), \qquad (19)$$

where $\Delta_{\mathcal{M}}(k, l) = \|I_{2,k}(l) - I_{1,k+\mathbf{m}^h(k,l)}(l + \mathbf{m}^v(k, l))\|_2^2$ represents the error of matching the pixel at position (k, l) in the second image with a pixel in the first image that is selected according to the motion information. As discussed in Section 3.2, the effect of motion between images can be captured by a linear operator A that is a composition of sub-matrices A^k. We can therefore rewrite Eq. (19) as

ARTICLE IN PRESS

V. Thirumalai, P. Frossard/J. Vis. Commun. Image R. xxx (2012) xxx–xxx 5

$$\widetilde{E}_d(\mathcal{M}) = \sum_{k=1}^{N_1} \|I_{2,k}^T - A_{\mathcal{M}}^k I_1\|_2^2, \tag{20}$$

where the sub-matrix A^k depends on the motion field $\mathcal{M}$ according to Eq. (6). In the rest of the development, we drop the index $\mathcal{M}$ as the dependency on the motion field is clear from the context.

In our framework however we do not have access to the original images, but only to the measurement vectors Y_1 and Y_2. We thus approximate the data cost $\widetilde{E}_d(\mathcal{M})$ by $E_d(\mathcal{M})$ that is computed directly from the measurement vectors. It can be written as

$$\widetilde{E}_d(\mathcal{M}) = \sum_{k=1}^{N_1} \left\| I_{2,k}^T - A^k I_1 \right\|_2^2 \tag{21}$$

$$\approx \sum_{k=1}^{N_1} \|Y_{2,k} - B^k Y_1\|_2^2 \tag{22}$$

$$\approx \sum_{k=1}^{N_1} \|Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1\|_2^2 \tag{23}$$

$$= E_d(\mathcal{M}). \tag{24}$$

Note that the data cost approximation due to working in the compressed domain comes from the relation between the matrices A and B that is given as $B \approx \Phi_2 A \Phi_1^T$. We study in details the effect of this approximation in the next section.

We can finally rewrite the regularized energy objective function for the correlation estimation problem. It reads as

$$E(\mathcal{M}) = \sum_{k=1}^{N_1} \left\| Y_{2,k} - \phi_2^k A_{\mathcal{M}}^k \Phi_1^T Y_1 \right\|_2^2 + \lambda \sum_{\mathbf{z},\mathbf{z}' \in \mathcal{N}} V_{\mathcal{M}}(\mathbf{z},\mathbf{z}'). \tag{25}$$

This cost function is used in the optimization problem of Eq. (16), which becomes a non-convex problem. The search space is discrete and usually constrained by limits on each of the motion values which typically define a motion search window. The solution to this problem can be determined with strong optimization techniques based on Graph Cuts [24,26] or Belief Propagation [27]. A comprehensive overview of various energy minimization techniques is summarized in [28]. In this paper, we use an optimization algorithm based on α-expansion mode in Graph Cuts whose complexity is bounded by a low order polynomial [26].

Finally, it should be noted that the correlation estimation can also be performed by block of pixels. In this case, each block of pixels is assumed to move in a coherent way, and the objective of the correlation estimation problem is to compute one motion vector per block. The data cost function can then be modified in a straightforward way by imposing the same motion vector for all the pixels in a block. The smoothness function is also modified in this case such that it penalizes the difference between the motion values of adjacent blocks rather than neighboring pixels. The optimization problem keeps the same form in block-based motion estimation but the search space is dramatically reduced as the number of motion vectors is smaller.

4.2. Compressed domain penalty

We now discuss the penalty of estimating the correlation from measurements instead of original images. When the correlation $\mathcal{M}$ is given, this penalty corresponds to the difference between the values of the regularized energy function of Eq. (16) that is evaluated from the original images and the respective linear measurements. Recall that the smoothness cost function $E_s(\mathcal{M})$ depends only on the correlation (see Eq. (17)). Therefore, the estimation penalty is identical to the error between the data cost functions $\widetilde{E}_d(\mathcal{M})$ and $E_d(\mathcal{M})$ that are computed in the original and compressed domains respectively. In what follows, we first show that

the penalty is bounded. Then, we show that the penalty decreases monotonically when the number of measurements increases.

Proposition 1. *The penalty of estimating the correlation from measurements is bounded. In particular we have* $|(1-\delta)^2 \widetilde{E}_d(\mathcal{M}) - C_l| \leqslant E_d(\mathcal{M}) \leqslant (1+\delta)^2 \widetilde{E}_d(\mathcal{M}) + C_u$, *where* $\delta > 0$, $C_l = \eta^2 + 2(1-\delta)\alpha\eta$, $C_u = \eta^2 + 2(1+\delta)\alpha\eta$, $\alpha = \sum_{k=1}^{N_1}\|I_{2,k}^T - A^k I_1\|_2$, $\eta = \sum_{k=1}^{N_1} \sigma_{max}(A^k)\sum_{p=k-w_y}^{k+w_y}\|\widetilde{I}_{1,p} - I_{1,p}\|_2$ *and* $\widetilde{I}_1 = \Phi_1^T Y_1$.

Proof 1. Let us assume that $\mathcal{M}$ or equivalently A is given. Then, the points $\mathcal{P} = \left\{ I_{2,k}^T, A^k I_1 : k = 1, 2, \ldots, N_1 \right\}$ forms a finite set. According to the Johnson–Lindenstrauss (JL) lemma, the distances between points in $\mathcal{P}$ are preserved in the measurement domain $\mathbb{R}^M$ when $M = \mathcal{O}(\delta^{-2}log|\mathcal{P}|)$ measurements are computed with a measurement matrix ϕ_2^k [29,30], where $|\mathcal{P}|$ denotes the number of points in $\mathcal{P}$. Mathematically, the JL-embedding is given as

$$(1-\delta)\left\| I_{2,k}^T - A^k I_1 \right\|_2 \leqslant \left\| \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 \right\|_2 \leqslant (1+\delta)\|I_{2,k}^T - A^k I_1\|_2, \tag{26}$$

for a positive constant δ. It should be noted that, when the measurement matrix ϕ_2^k satisfies Eq. (26), then with high probability it satisfies the restricted isometry property (RIP). For more details related to the connection between the JL-lemma and the RIP we refer the reader to [29,30]. Eq. (26) holds with high probability not only for measurement matrices constructed using Gaussian and Bernoulli distributions but also for structured measurement matrices constructed using orthonormal bases, e.g., DCT, FFT [30]. In our experiments we construct measurement matrices using structured FFT.

For a given row index k, the term $Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1$ in Eq. (23) we can write as

$$\begin{aligned} Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1 &= \phi_2^k I_{2,k}^T - \phi_2^k A^k \Phi_1^T \Phi_1 I_1 \\ &= \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 + \phi_2^k A^k I_1 - \phi_2^k A^k \Phi_1^T \Phi_1 I_1 \\ &= \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 + E I_1, \end{aligned} \tag{27}$$

where $E = \phi_2^k A^k - \phi_2^k A^k \Phi_1^T \Phi_1$. The term $\|Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1\|_2$ can be upper bounded as

$$\begin{aligned} \|Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1\|_2 &= \|\phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 + E I_1\|_2 \\ &\leqslant \left\| \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 \right\|_2 + \|E I_1\|_2 \\ &\leqslant (1+\delta)\left\| I_{2,k}^T - A^k I_1 \right\|_2 + \|E I_1\|_2, \end{aligned} \tag{28}$$

where the last inequality is derived from Eq. (26). Similarly the term $\left\| Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1 \right\|_2$ can be lower bounded as

$$\begin{aligned} \left\| Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1 \right\|_2 &= \left\| \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 + E I_1 \right\|_2 \\ &= \left\| \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 - (-E I_1) \right\|_2 \\ &\geqslant \left| \left\| \phi_2^k I_{2,k}^T - \phi_2^k A^k I_1 \right\|_2 - \|E I_1\|_2 \right| \tag{29} \\ &\geqslant |(1-\delta)\left\| I_{2,k}^T - A^k I_1 \right\|_2 - \|E I_1\|_2|, \tag{30} \end{aligned}$$

where Eq. (29) follows from $\|x - y\|_2 \geqslant |\|x\|_2 - \|y\|_2|$, and Eq. (30) is derived from Eq. (26). The term $\|E I_1\|_2$ in Eqs. (28) and (30) can also be bounded as

$$\|EI_1\|_2 = \left\| \phi_2^k A^k I_1 - \phi_2^k A^k \Phi_1^T \Phi_1 I_1 \right\|_2$$

$$= \left\| \phi_2^k A^k (\Phi_1^T \Phi_1 I_1 - I_1) \right\|_2$$ (31)

$$\leqslant \| A^k (\widetilde{I}_1 - I_1) \|_2$$

$$\leqslant \| A^k \|_2 \sum_{p=k-w_y}^{k+w_y} \| \widetilde{I}_{1,p} - I_{1,p} \|_2$$

$$= \sigma_{max}(A^k) \sum_{p=k-w_y}^{k+w_y} \| \widetilde{I}_{1,p} - I_{1,p} \|_2 = \eta_k,$$ (32)

where Eq. (31) follows from $\|\Phi x\|_2 \leqslant \|x\|_2$, as Φ is a non expanding operator [29] and $\widetilde{I}_1 = \Phi_1^T \Phi_1 I_1$ is the pre-image of I_1. $\sigma_{max}(A^k)$ in Eq. (32) denotes the largest singular value of A^k. The summation in Eq. (32) is carried out from rows $k - w_y$ to $k + w_y$ as the search window is usually bounded, where w_y is the admissible search size along the vertical direction. Combining Eqs. (28), (30) and (32), and by taking squares we get for each row of pixels

$$\left(|(1-\delta)\| I_{2,k}^T - A^k I_1 \|_2 - \eta_k| \right)^2 \leqslant \left\| Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1 \right\|_2^2$$

$$\leqslant \left((1+\delta)\| I_{2,k}^T - A^k I_1 \|_2 + \eta_k \right)^2.$$ (33)

Adding the second and third inequality terms of Eq. (33) for all rows $k = 1, 2, \ldots, N_1$ results in

$$E_d(\mathcal{M}) = \sum_{k=1}^{N_1} \left\| Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1 \right\|_2^2$$

$$\leqslant (1+\delta)^2 \sum_{k=1}^{N_1} \left\| I_{2,k}^T - A^k I_1 \right\|_2^2 + \sum_{k=1}^{N_1} \eta_k^2$$

$$+ \sum_{k=1}^{N_1} 2(1+\delta)\eta_k \left\| I_{2,k}^T - A^k I_1 \right\|_2$$

$$\leqslant (1+\delta)^2 \widetilde{E}_d(\mathcal{M}) + \left(\sum_{k=1}^{N_1} \eta_k \right)^2$$

$$+ 2(1+\delta) \underbrace{\sum_{k=1}^{N_1} \eta_k}_{\eta} \underbrace{\sum_{k=1}^{N_1} \left\| I_{2,k}^T - A^k I_1 \right\|_2}_{\alpha}$$

$$= (1+\delta)^2 \widetilde{E}_d(\mathcal{M}) + \eta^2 + 2(1+\delta)\eta\alpha$$

$$= (1+\delta)^2 \widetilde{E}_d(\mathcal{M}) + C_u,$$ (34)

where

$$C_u = \eta^2 + 2(1+\delta)\eta\alpha$$ (35)

and

$$\eta = \sum_{k=1}^{N_1} \eta_k = \sum_{k=1}^{N_1} \sigma_{max}(A^k) \sum_{p=k-w_y}^{k+w_y} \| \widetilde{I}_{1,p} - I_{1,p} \|_2.$$ (36)

In a similar way, from the first and second inequality terms of Eq. (33) we get

$$E_d(\mathcal{M}) = \sum_{k=1}^{N_1} \left\| Y_{2,k} - \phi_2^k A^k \Phi_1^T Y_1 \right\|_2^2$$

$$\geqslant \sum_{k=1}^{N_1} \left\{ (1-\delta)^2 \left\| I_{2,k}^T - A^k I_1 \right\|_2^2 + \eta_k^2 - 2(1-\delta)\eta_k \left\| I_{2,k}^T - A^k I_1 \right\|_2 \right\}$$

$$\geqslant \left| \sum_{k=1}^{N_1} (1-\delta)^2 \left\| I_{2,k}^T - A^k I_1 \right\|_2^2 + \sum_{k=1}^{N_1} \eta_k^2 - \sum_{k=1}^{N_1} 2(1-\delta)\eta_k \left\| I_{2,k}^T - A^k I_1 \right\|_2 \right|$$

$$\geqslant \left| (1-\delta)^2 \sum_{k=1}^{N_1} \| I_{2,k}^T - A^k I_1 \|_2^2 - \left(\sum_{k=1}^{N_1} \eta_k \right)^2 - 2(1-\delta) \sum_{k=1}^{N_1} \eta_k \sum_{k=1}^{N_1} \| I_{2,k}^T - A^k I_1 \|_2 \right|$$

$$= |(1-\delta)^2 \widetilde{E}_d(\mathcal{M}) - (\eta^2 + 2(1-\delta)\eta\alpha)|$$

$$= |(1-\delta)^2 \widetilde{E}_d(\mathcal{M}) - C_l|,$$ (37)

where

$$C_l = \eta^2 + 2(1-\delta)\eta\alpha$$ (38)

and η is given in Eq. (36). □

Proposition 2. *The penalty of estimating the correlation from measurements monotonically decreases when the measurement rate K/N increases. It further becomes negligible at high measurement rate.*

Proof 2. In Proposition 1 we have shown that the difference between the data cost functions estimated from compressed measurements $E_d(\mathcal{M})$ and images $\widetilde{E}_d(\mathcal{M})$ is lower and upper bounded by errors C_l and C_u, respectively given in Eq. (38) and Eq. (35). The error $\eta = \sum_{k=1}^{N_1} \eta_k \propto \sum_{k=1}^{N_1} \| \widetilde{I}_{1,k} - I_{1,k} \|_2$ (see Eq. (36)) decreases with increasing measurement rate because $(\phi_1^k)^T \phi_1^k$ is an orthogonal projection operator. Eventually, $\widetilde{I}_1 = \Phi_1^T \Phi_1 I_1$ becomes arbitrarily close to I_1 when the number of measurements increases. Therefore, the errors C_l and C_u decrease as the measurement rate increases and when sufficient number of measurements are taken the errors C_l and C_u become negligible, i.e., $E_d(\mathcal{M}) \approx \widetilde{E}_d(\mathcal{M})$. □

Due to the error between the cost functions $E_d(\mathcal{M})$ and $\widetilde{E}_d(\mathcal{M})$, the solution $\mathcal{M}$ estimated from the linear measurements is not accurate especially at low measurement rates. The solution of the correlation estimation problem in the compressed domain might thus be quite far from the actual correlation between images. However, as the number of measurements increases the approximation in the compressed domain becomes more accurate and the solution of the correlation estimation tends to the actual correlation between original images.

5. Experimental results

5.1. Setup

We analyze the performance of the correlation estimation in both stereo and video imaging applications. The random projections are computed using a scrambled Fourier measurement matrix where the scrambled operator is a diagonal matrix with entries ±1 taken from an i.i.d. Bernoulli random variable with equal probability [30]. In all our experiments, we sample both the images using a same measurement rate. The correlation is estimated without prior reconstruction of images by minimizing the objective function in Eq. (25).

For the stereo imaging case, we evaluate the disparity estimation performance in two natural image sets namely *Tsukuba* and *Venus*[1] [31]. These datasets have been captured by a camera array where the different viewpoints correspond to translating the camera along one of the image coordinate axis. In such a scenario the motion of objects due to the viewpoint change is restricted to the horizontal direction with no motion along the vertical direction. The disparity estimation is thus a one-dimensional search problem, and the smoothness and data cost functions are modified accordingly by assuming that $\mathbf{m}^v = 0$. The size of the search windows used in our experiments are 16 pixels for Tsukuba and 20 pixels for Venus. In our experiments we estimate disparity in both dense (per pixel) and block settings, where the block size is fixed to 4×4 pixels.

In the video scenario, we analyze the motion estimation accuracy in two synthetic scenes, namely *Yosemite* and *Grove*, and one natural scene *Mequon*.[2] The Grove and Mequon datasets are resampled to a resolution of 160×120 pixels using bilinear filters.

---

[1] Available at http://vision.middlebury.edu/stereo/data/scenes2001/.
[2] Available at http://vision.middlebury.edu/flow.

ARTICLE IN PRESS

V. Thirumalai, P. Frossard/J. Vis. Commun. Image R. xxx (2012) xxx–xxx   7

The size of search windows is of ±3 pixels in both horizontal and vertical directions. For the sake of simplicity, we estimate a motion field for blocks of pixels with size of 4×4 pixels and not motion vectors for each pixel.

The accuracy of the correlation estimation is evaluated by comparing to the groundtruth information and to the correlation estimated from the reconstructed images. We propose another representation of the accuracy of the correlation by discussing the quality of the second view $\widehat{I}_2$ that is reconstructed by prediction of the first frame according to the estimated correlation. We then analyze the influence of the sampling matrix and the effect of measurement quantization on the correlation estimation performance. We finally show the importance of accurate correlation estimation in a novel joint reconstruction algorithm where images are reconstructed from measurements while satisfying sparsity constraints as well as consistency with the correlation information. Note that in practice, the groundtruth correlation model and the original images are not available a priori to estimate an optimal regularization parameter λ. In such cases, the regularization parameter λ can be estimated based on learning from a set of training images or using the automated method proposed in [32]. In our experiments, we however select the parameter λ based on trial and error experiments.

5.2. Disparity estimation performance

We first illustrate disparity maps for the Venus dataset where the compressed data have been obtained with a different measurement matrix for each image, i.e., $\Phi_1 \neq \Phi_2$. Fig. 3(b) and Fig. 3(d) show the disparity map from a measurement rate 0.2 and 0.7 respectively. Fig. 3(c) and Fig. 3(e) represent the corresponding disparity errors. Comparing the results with the groundtruth given in Fig. 3(a) we see that at low measurement rate (corresponding to 0.2) we estimate a coarse version of the disparity map. Quantitatively, the disparity error with respect to the groundtruth is found

out to be 41% when measured as the percentage of pixels where the absolute error is greater than one [31] as shown in Fig. 3(c). At higher rate, the disparity map is more accurate and the disparity error drops below 11%.

We then show the quality of the reconstruction of the second image that is predicted from the first image using the correlation estimate. When a coarse disparity map $\mathbf{m}^h$ (i.e., estimated at low measurement rate) is used for image prediction the resulting predicted image $\widehat{I}_2$ is closer to I_2 than I_1 (see Fig. 4(a) and Fig. 4(b) respectively). We observe that the mean square error (MSE) between the predicted image $\widehat{I}_2$ and I_2 is smaller than the error between $\widehat{I}_2$ and I_1, which confirms the benefit of the disparity estimate in the prediction. When the measurement rate increases the quality of the disparity map improves and the quality of the predicted image $\widehat{I}_2$ also improves substantially as it can be observed in Fig. 4(c) and Fig. 4(d) where the measurement rate is set to 0.7.

The next experiments illustrate the benefit of regularization in the disparity estimation problem. Fig. 5 plots the quality of the predicted image $\widehat{I}_2$ with and without smoothness cost (i.e., $\lambda \neq 0$ and $\lambda = 0$ respectively in Eq. (25) respectively) for the Tsukuba dataset. It is clear that the quality of $\widehat{I}_2$ is improved by enabling the regularization term in our optimization framework. Similar experimental finding is observed for the Venus dataset. From Fig. 5 we further observe that the quality of the predicted image $\widehat{I}_2$ at a measurement rate 0.05 is 22.2 dB (the corresponding disparity error is 39%), which is approximately 3.5 dB away from the saturation point or from global minima solution due to influence of the penalty terms C_l and C_u discussed in Section 4.2. As the measurement rate increases, the influence of the terms C_l and C_u decreases. As a result, the quality of predicted image $\widehat{I}_2$ increases with the measurement rate and saturates with measurement rates above 0.5. In other words, our scheme gives optimal disparity solution at high measurement rate. We also carry out experiments using the same measurement matrix for both images,
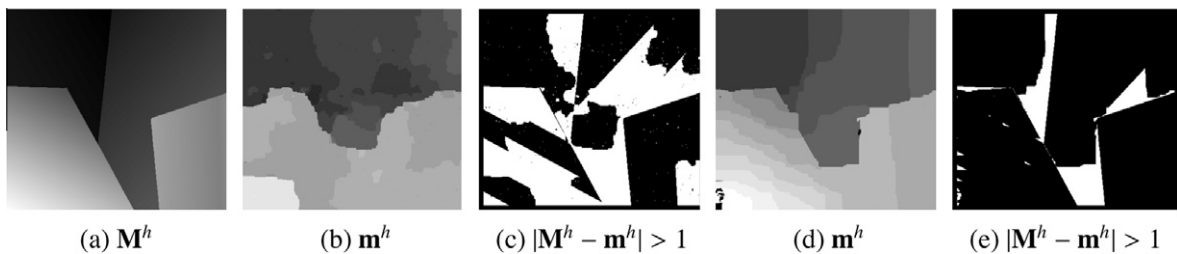

(a) $\mathbf{M}^h$ (b) $\mathbf{m}^h$ (c) $|\mathbf{M}^h - \mathbf{m}^h| > 1$ (d) $\mathbf{m}^h$ (e) $|\mathbf{M}^h - \mathbf{m}^h| > 1$

Fig. 3. Comparison of the estimated disparity image with respect to groundtruth information at measurement rates 0.2 and 0.7 in the Venus dataset. (a) Groundtruth disparity image $\mathbf{M}^h$; (b) computed dense disparity image $\mathbf{m}^h$ at measurement rate 0.2; (c) disparity error at rate 0.2. The pixels with absolute error greater than one is marked in white. The percentage of white pixels is 41%. (d) Computed dense disparity image $\mathbf{m}^h$ at measurement rate 0.7; (e) disparity error at rate 0.7. The percentage of white pixels is 10.7%.

(a) MSE: 205   (b) MSE: 1221   (c) MSE: 101   (d) MSE: 1219

Fig. 4. Evaluating the accuracy of disparity image in Fig. 3(b) and Fig. 3(d) in terms of image prediction quality for the Venus dataset. The disparity images in Fig. 3(b) and Fig. 3(d) are used to predict the image $\widehat{I}_2$ at measurement rates 0.2 and 0.7 respectively. (a) Inverse prediction $1 - |\widehat{I}_2 - I_2|$ at a measurement rate of 0.2; (b) inverse prediction $1 - |\widehat{I}_2 - I_1|$ at a measurement rate of 0.2; (c) inverse prediction $1 - |\widehat{I}_2 - I_2|$ at a measurement rate of 0.7; (d) inverse prediction $1 - |\widehat{I}_2 - I_1|$ at a measurement rate of 0.7. The error is inverted, so that the white pixels correspond to no error.
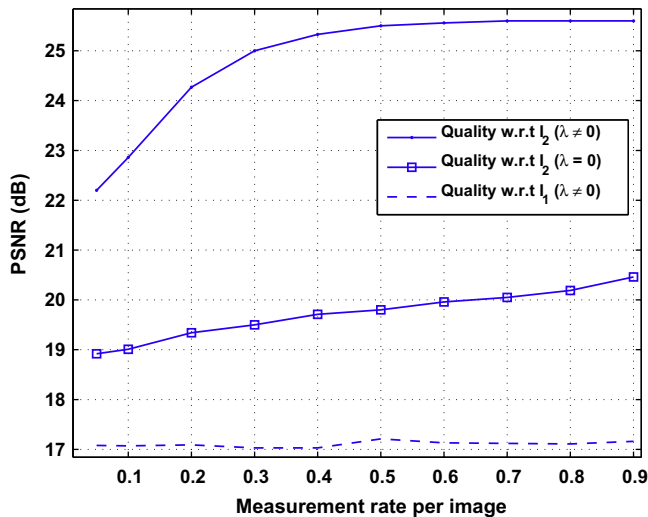
Fig. 5. Comparison of the quality of predicted image $\widehat{I}_2$ with respect to I_2 and I_1 with and without regularization, i.e., $\lambda \neq 0$ and $\lambda = 0$ in Eq. (25) respectively in the Tsukuba dataset. The image prediction is carried out using dense disparity image.

i.e., $\Phi_1 = \Phi_2$. Fig. 6(a) and Fig. 6(b) compare the PSNR quality of the predicted image $\widehat{I}_2$ and the disparity error (DE), respectively with the results obtained with different measurement matrices. It is clear that the prediction image quality and the disparity accuracy improve when different measurement matrices are used, since this brings more information from both images to solve the correspondence problem.

We finally compare our disparity estimation results to a scheme that first reconstructs the images before estimating the disparity map. The images are reconstructed independently from the corresponding measurements by solving a convex optimization problem. We denote this methodology as *disparity from reconstructed images* (DFR). We have tried out two different reconstruction methodologies: (1) DFR-sparsity that minimizes the l_1 norm of the sparse coefficients assuming that the image is sparse in a particular orthonormal basis (e.g., a wavelet basis); this problem is solved using GPSR [33]; (2) DFR–TV that minimizes the TV norm of the reconstructed image; this problem is solved using BPDQ toolbox [34].
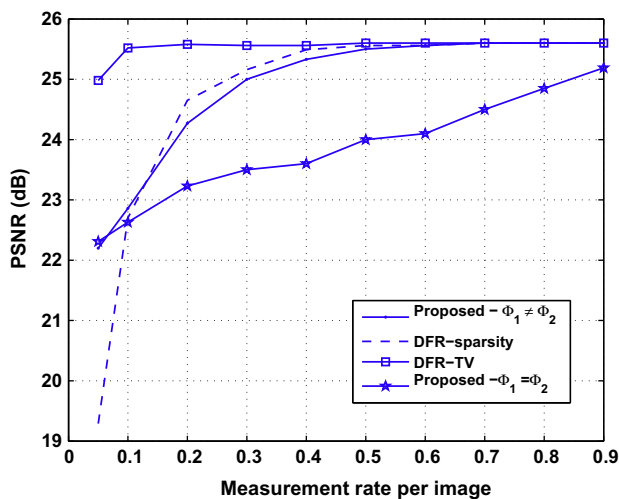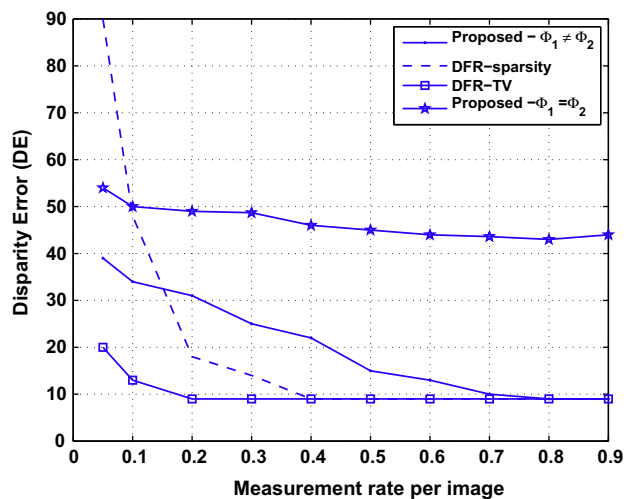
The disparity map is then estimated using α-expansion mode in Graph Cuts applied on the reconstructed images. Fig. 6 shows the comparison of the proposed scheme with the DFR-sparsity and DFR–TV schemes for the Tsukuba dataset. From Fig. 6(a) and Fig. 6(b) we observe that the performance of our low complexity correlation solution competes with DFR-sparsity scheme. Especially at rates smaller than 0.1, our scheme performs better than the DFR-sparsity scheme as the poor image reconstruction quality in the DFR-sparsity scheme leads to a bad estimation of disparity map. On the other hand, DFR–TV scheme significantly outperforms our scheme at lower rates due to the good quality of reconstructed images. However, our scheme estimates the optimal correlation for rates above 0.5 and thus performs similar to the DFR–TV scheme at high measurement rate but with a complexity that is dramatically smaller as it avoids image reconstruction. In particular, in our experiments we have observed that the running time of Graph Cuts algorithms that estimate the correlation information from linear measurements is approximately same as the one that estimates the correlation information from reconstructed images. The complexity of our correlation estimation scheme stays reasonable due to the efficiency of Graph Cuts algorithms whose complexity is bounded by a low order polynomial [26]. Comparing to the DFR-sparsity and DFR–TV schemes, we save on the complexity corresponding to solving the l_2–l_1 and l_2–TV optimization problems respectively. It is however hard to precisely give the order of complexity of solving the l_2–l_1 and l_2–TV optimization problems, as it is highly depend on the type of solvers. For some of the popular solvers like GPSR [33] and NESTA [35], the order of complexity is given as $T\mathcal{O}(N\log N)$, where T is the number of iterations and N is the resolution of the image [36]. Therefore, comparing to the DFR schemes we save a complexity of $2T\mathcal{O}(N\log N)$.

The complexity of the proposed scheme can be further reduced when a disparity value is estimated per block instead of per pixel. Fig. 7 compares the performance of our scheme with respect to the DFR schemes when disparity is estimated using per block with block size 4×4. Comparing Fig. 7 and Fig. 6(a) we see that the relative performances between the schemes remains approximately the same when the disparity image is estimated per block or per pixel. This confirms that the proposed scheme can easily adapt the granularity of disparity estimation without big penalty in order to meet the complexity requirements at the decoder.
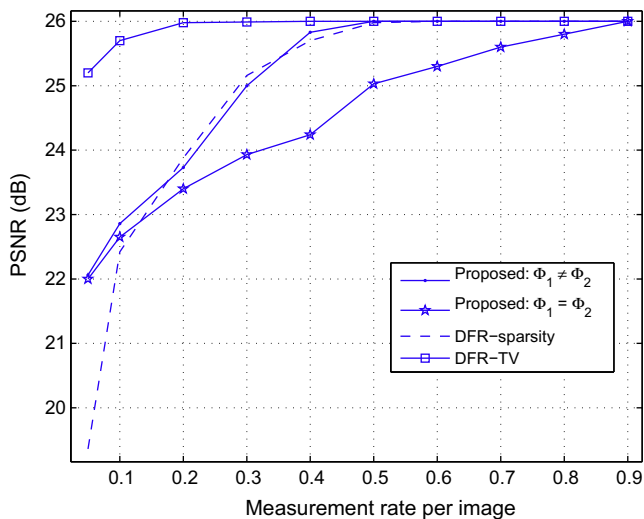

(a)

(b)

Fig. 6. Comparison of the proposed scheme with the DFR schemes in the Tsukuba dataset: (a) comparison in terms of image prediction quality; (b) comparison in terms of disparity error. The performance of the proposed scheme is evaluated using both the same and different sets of measurement matrices.

Fig. 7. Comparison of the proposed scheme over disparity from reconstructed image (DFR) schemes for the Tsukuba dataset. The disparity is estimated per block with a block size 4×4. The performance of the proposed scheme is evaluated using both the same and different sets of measurement matrices.

5.3. Motion estimation performance

We now illustrate the performance of our correlation estimation algorithm in video sequences. We first estimate motion vectors per blocks of 4×4 pixels with different sensing matrices $\Phi_1 \neq \Phi_2$ for each image. These vectors are then used to predict the second image from the first image. Fig. 8 compares the predicted image $\widehat{I}_2$ with the original images I_2 and I_1 for the Yosemite and Grove datasets respectively. It is clear that for a given measurement rate the predicted image $\widehat{I}_2$ is closer to I_2 than I_1 which indicates that the motion between images is efficiently captured by our correlation estimation algorithm. Similar experimental results are observed in the Mequon dataset.

We then highlight the benefit of sampling the images with different sets of measurement matrices in Fig. 9. From Fig. 9 we see that the quality of the predicted image $\widehat{I}_2$ is better when the images are sampled with different measurement matrices, compared to the case where the same sampling matrix is used for all images.
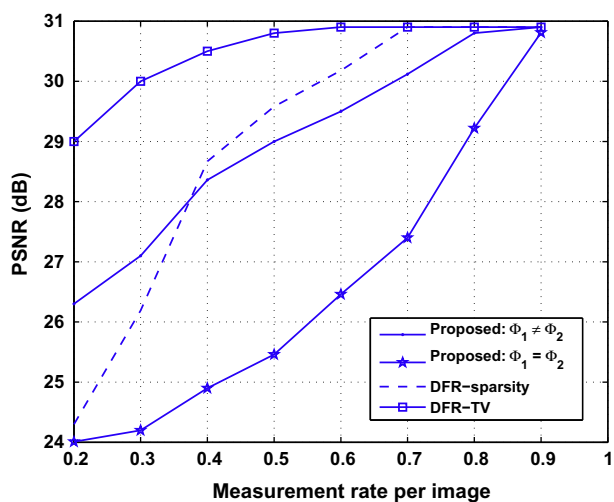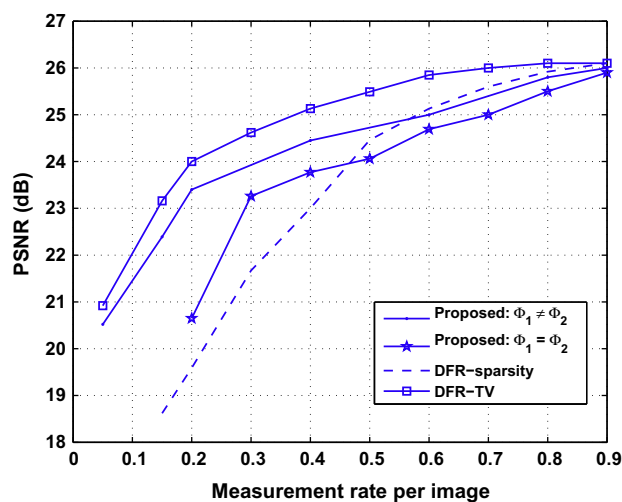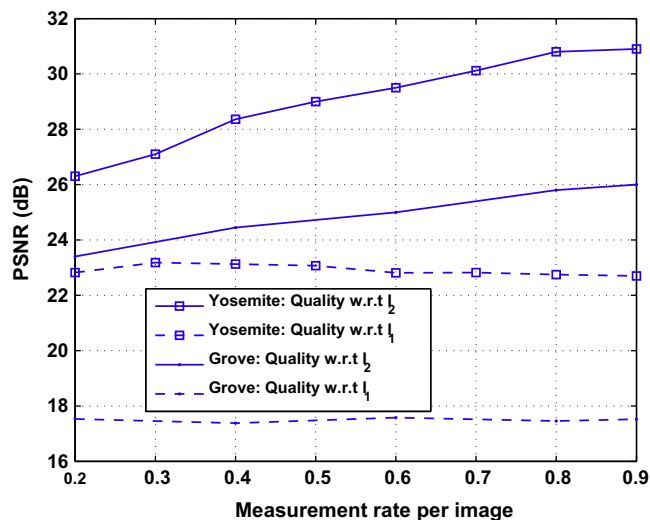

Fig. 8. Illustration of the accuracy of motion field estimation in terms of image prediction quality for the Yosemite and Grove datasets. The quality of the predicted image $\widehat{I}_2$ is compared with respect to I_2 and I_1. The prediction is carried out using the motion field estimated with block of pixels 4×4.

This confirms the results shown for the disparity estimation performance. We finally compare our results to the DFR-sparsity and DFR–TV schemes that build the correlation model from (independently) reconstructed images based on minimizing the sparsity and TV priori respectively. Fig. 9(a) and Fig. 9(b) show the comparison for the Yosemite and Grove datasets respectively. From Fig. 9 we see that for both datasets the proposed scheme performs better than the DFR-sparsity scheme at low rates and competes with the DFR-sparsity scheme at high rates, as observed in the disparity estimation study. Similar experimental findings are observed in the Mequon dataset. Furthermore, we see that the proposed scheme is competitive with the performance of DFR–TV scheme at low rate for the Grove dataset (see Fig. 9(b)); as the Grove scene is textured with limited low frequency components, the *TV* prior in the reconstruction scheme results in poor reconstruction quality in the textured areas. Overall, the proposed scheme provides effective motion estimation results, while avoiding an order of computational complexity $2T\mathcal{O}(NlogN)$ involved in the image reconstruction steps with the DFR–TV and DFR-sparsity schemes.
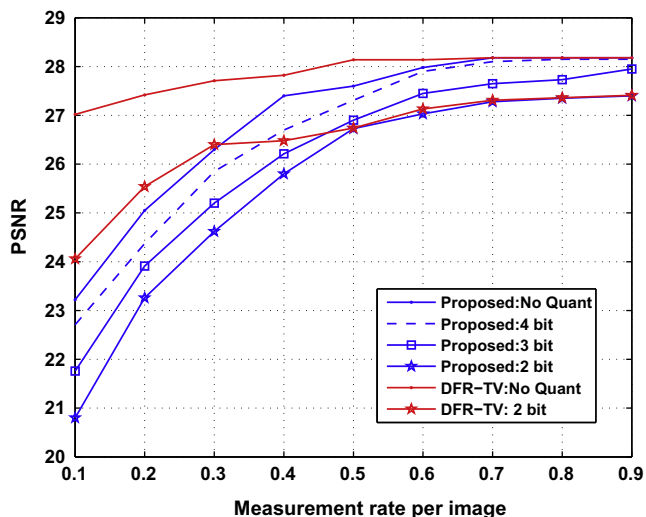

(a)

(b)

Fig. 9. Comparison of the quality of the predicted image $\widehat{I}_2$ between the proposed, DFR-sparsity and DFR–TV schemes: (a) Yosemite dataset; (b) Grove dataset. The image prediction is carried out using the motion field that is estimated using block of pixels 4×4.

ARTICLE IN PRESS

10   *V. Thirumalai, P. Frossard / J. Vis. Commun. Image R. xxx (2012) xxx–xxx*

Fig. 10. Effect of quantization on the quality of the predicted image $\widehat{I}_2$ in the Venus dataset. When the measurements are quantized we used $BPDN_p$ [37] to reconstruct the images in DFR–TV based scheme. The image prediction is carried out using the dense disparity image estimated by solving Eq. (25).

5.4. Measurements quantization

We briefly study here the performance of the correlation estimation algorithm when the measurements are affected by noise and in particular quantization noise. For the sake of simplicity we quantize the measurements using a uniform quantizer and we denote the quantized measurements as $\widehat{Y}_1$ and $\widehat{Y}_2$. We then estimate the correlation model by minimizing the energy in Eq. (25) using the quantized measurements $\widehat{Y}_1$ and $\widehat{Y}_2$.

Fig. 10 shows the effect of quantization on the disparity estimation performance when measurements are uniformly quantized using 2-, 3- and 4-bits. We use the disparity map to predict the second image in the Venus dataset. Interestingly we see that the 4-bit quantizer does not significantly affect the quality of the disparity image, as the degradation hardly reaches 0.5 dB in the quality of $\widehat{I}_2$ at low to medium rates. As expected however the quality of the predicted image $\widehat{I}_2$ decreases with increasing quantizer coarse-

ness level for a fixed measurement rate. We then compare our results with the DFR–TV scheme that reconstructs the images by solving an optimization problem based on $BPDN_p$ in order to efficiently handle the quantization noise [37] and then use the reconstructed images for disparity estimation. From Fig. 10 we see that the performance gap between the DFR–TV and compressed domain estimation is approximately the same in both the unquantized and quantized (i.e., 2-bit) scenarios. However, it should be noted that the DFR–TV scheme considers the nonlinearities due to quantization while reconstructing the images. Such effects are not considered in the proposed scheme. The solution of our scheme could also be improved by considering the quantization non-linearities but this problem is beyond the scope of this paper.

We also study the effect of measurement quantization on the quality of the motion field. Fig. 11(a) and Fig. 11(b) shows the quality of the predicted image where the prediction is performed with motion vectors estimated from quantized measurements. The quality of the predicted image or equivalently the accuracy of motion estimation is reduced when the measurements are quantized as expected. Similarly to the case of disparity estimation, the influence of quantization is negligible when the measurements are quantized with a 4-bit quantizer. We also compare our results to the DFR–TV scheme that reconstructs the images by solving an optimization problem based on $BPDN_p$ and estimate motion from the reconstructed images. The performance of the DFR–TV scheme for the 2-bit quantization scenario is shown in Fig. 11. Interestingly, when the measurements are quantized we see that the proposed scheme competes with the DFR–TV scheme; this is because of the poor image reconstruction performance in the DFR–TV scheme when the measurements are coarsely quantized (i.e., 2-bit quantizer).

5.5. Importance of correlation in joint reconstruction

We finally propose to study the importance of accurate correlation estimation in a novel joint reconstruction algorithm (see Fig. 1). We propose to reconstruct a pair of images $\acute{I}_1$ and $\acute{I}_2$ by enforcing consistency with the compressed information and also with the estimated correlation model. A pair of image $\acute{I}_1$ and $\acute{I}_2$ is reconstructed as a solution of the following constrained optimization problem:
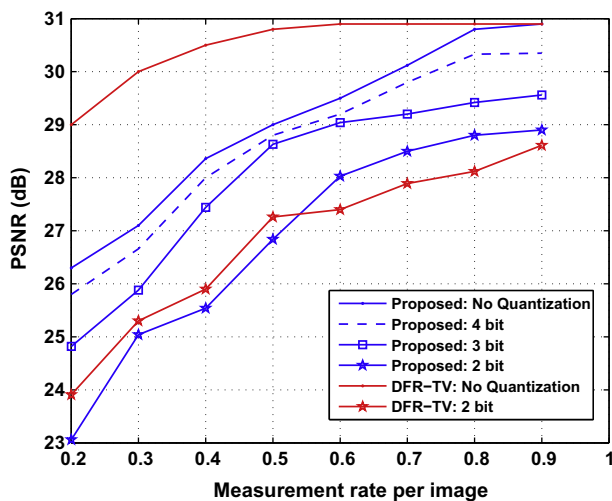
(a)
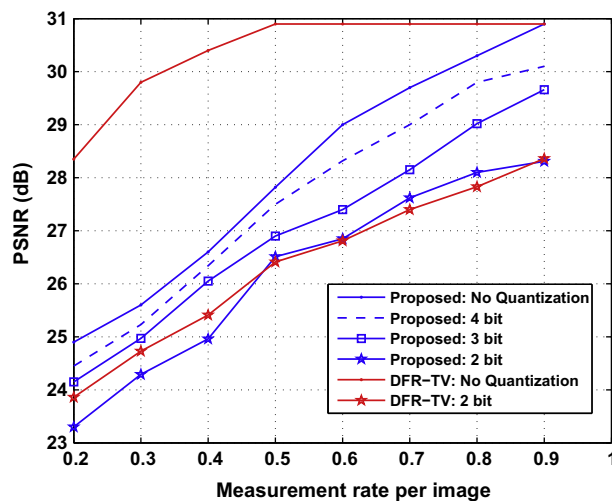
(b)

Fig. 11. Effect of using 4-, 3- and 2-bits quantizers on the image prediction quality for (a) Yosemite dataset and (b) Mequon dataset. When the measurements are quantized we used $BPDN_p$ [37] to reconstruct the images in DFR–TV scheme. The image prediction is carried out using the motion field estimated by solving Eq. (25)
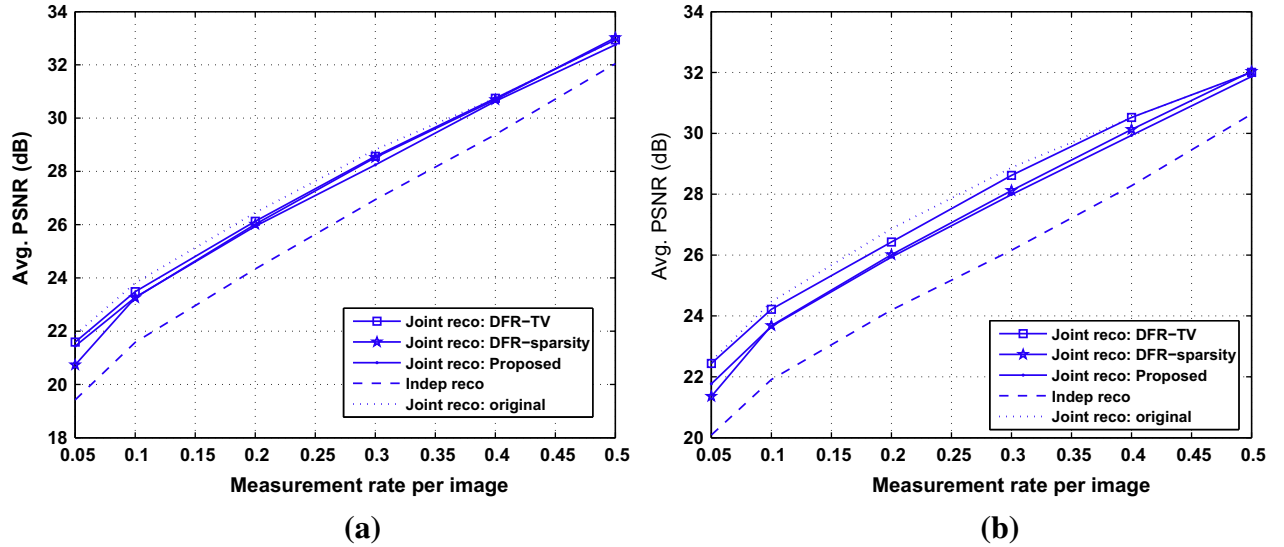
ARTICLE IN PRESS

V. Thirumalai, P. Frossard / J. Vis. Commun. Image R. xxx (2012) xxx–xxx
11

Fig. 12. Influence of disparity accuracy on the joint reconstruction performance for (a) Tsukuba and (b) Venus datasets. The joint reconstruction is carried out using the dense disparity image in the proposed and DFR-based schemes. The joint reconstruction performance is also compared to the independent reconstruction scheme.

$$(\acute{I}_1, \acute{I}_2) = arg \min_{I_1, I_2}(\|\psi^* I_1\|_1 + \|\psi^* I_2\|_1) \quad \text{s.t. } \|Y_1 - \Phi_1 I_1\|_2 = 0,$$

$$\|Y_2 - \Phi_2 I_2\|_2 = 0, \quad \|I_2 - A I_1\|_2^2 \leqslant \epsilon, \tag{39}$$

where ψ is a redundant dictionary or an orthonormal basis in which the image is assumed to be sparse and ψ^* is the conjugate transpose of ψ. From Eq. (39) it is clear that the images can be reconstructed independently if we solve the optimization without the last constraint $\|I_2 - A I_1\|_2^2 \leqslant \epsilon$. This corresponds to the independent reconstruction of sparse images in ψ which agrees with the measurement information. By adding the last constraint, we impose that the pair of images also fit with the correlation model, in addition to the sparsity and data fidelity constraints. As a result the reconstruction quality for a given measurement rate is better when the images are reconstructed jointly than independently. The optimization problem for joint reconstruction can be re-written with proximity operators and solved efficiently using the parallel proximal algorithm (PPXA) proposed by Combettes et al. [38] (see [39] for a similar solution in an asymmetric joint decoding scheme).

We analyze the performance of our joint reconstruction scheme with a constraint imposed by the correlation estimated from linear measurements. In particular, we perform joint reconstruction experiments using the correlation model that is estimated using a different measurement matrix for each image, i.e., $\Phi_1 \neq \Phi_2$. We assume that the image is sparse in an orthonormal basis constructed using a wavelet transform. In our experiments, the parameter ϵ in the optimization problem is selected based on a trial and error procedure that maximizes the reconstruction image quality $\acute{I}_1$ and $\acute{I}_2$, and we set $\epsilon = 14$. We first compare our results to an independent reconstruction scheme that does not exploit the correlation between the images (i.e., the constraint $\|I_2 - A I_1\|_2^2 \leqslant \epsilon$ is removed in Eq. (39)). Fig. 12 compares the average reconstruction quality between the joint (denoted as *Joint:Proposed*) and independent reconstruction schemes for the Tsukuba and Venus datasets. The joint reconstruction improves the reconstruction quality by 2 dB at low measurement rates and about 1 dB at high rates. We also observe in our experiments that the PSNR quality of the reconstructed images $\acute{I}_1$ and $\acute{I}_2$ is similar at a given measurement rate. The disparity estimate thus proves to be useful in improving the quality in the image reconstruction process.

We then jointly reconstruct the images using a disparity estimated with DFR-sparsity and DFR–TV schemes. Fig. 12 compares the quality of reconstructed images between the proposed and DFR schemes. We see that the disparity estimated from compressed measurements leads to a competitive performance with the disparity estimated by the DFR-sparsity scheme in terms of joint reconstruction quality. This is particularly obvious at low rate 0.05 where the DFR-sparsity scheme fails to accurately estimate the disparity. However, the quality of the reconstructed images is marginally penalized (i.e., 0.2 dB and 0.4 dB for the Tsukuba and Venus datasets respectively) compared to the reconstruction achieved when the disparity estimated by the DFR–TV scheme. Finally, we carry out the same experiments in a scenario where the images are jointly reconstructed using a correlation model that is estimated from the original images. This scheme serves as a benchmark for the joint reconstruction since the correlation is accurately known at the decoder. The corresponding results are denoted as *Joint reco:original* in Fig. 12. We see that the reconstruction quality achieved with the correlation estimated from compressed measurements converges to the performance benchmark when the measurement rate increases which further confirms the quality of the disparity estimation. Finally, it should be noted that similar tendencies have been observed in joint reconstruction of video frames, but the corresponding results are omitted here due to space constraints.

6. Conclusion

In this paper we have presented a framework for estimating the correlation between images given in the form of linear measurements without implementing explicit image reconstruction steps. We have proposed a linear representation of disparity and motion models and show that the correlation can be estimated in the compressed domain, thanks to the distance preserving property of the sensing matrix. The correlation is estimated by solving a regularized energy model that enforces consistency with the measurements and smoothness of the correlation information. Extensive experimental results demonstrate that the proposed methodology provides a good estimation of disparity or motion fields in different natural and synthetic image datasets, especially when images are

sampled with different measurement matrices. We also show that our correlation estimation solution competes with the correlation estimation solutions that reconstructs images a priori, but becomes clearly advantageous due to its lower computational complexity. The correlation estimation from compressed measurements thus provides an effective solution for distributed scene analysis or coding applications in low complexity sensor networks.

Acknowledgements

References

[1] V. Thirumalai, P. Frossard, Dense disparity estimation from linear measurements, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2011.

[2] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.

[3] E.J. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2006) 489–509.

[4] E.J. Candes, J. Romberg, Practical signal recovery from random projections, Proceedings of the SPIE Computational Imaging (2005).

[5] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, R. Baraniuk, Single-pixel imaging via compressive sampling, IEEE Signal Process. Magaz. 25 (2) (2008) 83–91.

[6] S. Mun, J. Fowler, Block compressed sensing of images using directional transforms, in: Proceedings of the IEEE International Conference on Image Processing, 2009.

[7] L. Gan, T.T. Do, T.D. Tran, Fast compressive imaging using scrambled Hadamard ensemble, in: Proceedings of the European Signal and Image Processing Conference, 2008.

[8] V. Stankovic, L. Stankovic, S. Cheng, Compressive video sampling, in: Proceedings of the European Signal and Image Processing Conference, 2008.

[9] J.Y. Park, M.B. Wakin, A multiscale framework for compressive sensing of video, in: Proceedings of the Picture Coding Symposium, 2009.

[10] N. Vaswani, Kalman filtered compressed sensing, in: Proceedings of the IEEE International Conference on Image Processing, 2008.

[11] S. Pudlewski, A. Prasanna, T. Melodia, Compressed sensing enabled video streaming for wireless multimedia sensor networks, IEEE Transactions on Mobile Computing (2012), doi:10.1109/TMC.2011.175.

[12] V.K. Goyal, A.K. Fletcher, S. Rangan, Compressive sampling and lossy compression, IEEE Signal Process. Magaz. 25 (2) (2008) 48–56.

[13] M.F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, Distributed compressed sensing of jointly sparse signals, in: Proceedings of the Asilomar Conference on Signal System and Computing, 2005.

[14] M.F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, R.G. Baraniuk, Universal distributed sensing via random projections, in: Proceedings of the Information Processing in Sensor Networks, 2006.

[15] L.W. Kang, C.S. Lu, Distributed compressive video sensing, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.

[16] T.T. Do, Y. Chen, D.T. Nguyen, N. Nguyen, L. Gan, T. Tran, Distributed compressed video sensing, in: Proceedings of the IEEE International Conference on Image Processing, 2009.

[17] J.P. Nebot, Y. Ma, T. Huang, Distributed video coding using compressive sampling, in: Proceedings of the Picture Coding Symposium, 2009.

[18] M. Trocan, T. Maugey, J.E. Fowler, B. Pesquet-Popescu, Disparity compensated compressed sensing reconstruction for multiview images, in: Proceedings of the IEEE International Conference on Multimedia and Expo, 2010.

[19] M. Trocan, T. Maugey, E.W. Tramel, J.E. Fowler, B. Pesquet-Popescu, Multistage compressed-sensing reconstruction of multiview images, in: Proceedings of the IEEE International workshop on Multimedia Signal Processing, 2010.

[20] V. Thirumalai, P. Frossard, Distributed representation of geometrically correlated images with compressed linear measurements, IEEE Transactions on Image Processing, URL <http://arxiv.org/abs/1111.5612>, submitted for publication.

[21] J.Y. Park, M.B. Wakin, A geometric approach to multi-view compressive imaging, EURASIP Journal on Advances in Signal Processing, in press.

[22] X. Li, Z. Wei, Z. Xiao, Compressed sensing joint reconstruction for multi-view images, Electron. Lett. 46 (23) (2010) 1548–1550.

[23] T. Do, T. Tran, L. Gan, Fast compressive sampling with structurally random matrices, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.

[24] O. Veksler, Efficient graph based energy minimization methods in computer vision, Ph.D. thesis, Cornell University, 1999.

[25] S. Baker, S. Roth, D. Scharstein, M. Black, J. Lewis, R. Szeliski, A database and evaluation methodology for optical flow, Int. J. Comput. Vis. 1 (92) (2011) 1–31. <http://www.springerlink.com/content/p516733117226378/> .

[26] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1222–1239.

[27] P. Felzenszwalb, D. Huttenlocher, Efficient belief propagation for early vision, Int. J. Comput. Vis. 70 (1) (2006) 41–54.

[28] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, C. Rother, A comparative study of energy minimization methods for markov random fields with smoothness-based priors, IEEE Trans. Pattern Anal. Mach. Intell. 30 (6) (2008) 1068–1080.

[29] R. Baraniuk, M. Davenport, R. DeVore, A simple proof of the restricted isometry property for random matrices, Constructive Approximation, vol. 28, Springer, 2008.

[30] T. Do, L. Gan, Y. Chen, N. Nguyen, T. Tran, Fast and efficient dimensionality reduction using Structurally Random Matrices, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.

[31] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense stereo, Int. J. Comput. Vis. 47 (2002) 7–42.

[32] L. Zhang, S.M. Seitz, Estimating optimal parameters for MRF stereo from a single image pair, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 331–342.

[33] M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, IEEE J. Selected Topics Signal Process. 1 (2007) 586–597.

[34] D.K. Hammond, L. Jacques, M. Fadili, G. Puy, P. Vandergheynst, The Basis Pursuit DeQuantizer (BPDQ) toolbox, URL <http://wiki.epfl.ch/bpdq>, 2009.

[35] S. Becker, J. Bobin, E. Candes, NESTA: a fast and accurate first-order method for sparse recovery, Tech. Rep. Caltech (2009).

[36] Y.L. Montagner, E. Angelini, J.-C. Olivo-Marin, Comparison of reconstruction algorithms in compressed sensing applied to biological imaging, in: Proceedings of the IEEE International Symposium on Biological Imaging, 2011.

[37] L. Jacques, D.K. Hammond, M.J. Fadili, Dequantizing compressed sensing: when oversampling and non-gaussian constraints combine, IEEE Trans. Inform. Theory 57 (2011) 559–571.

[38] P.L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in: Fixed point algorithms for inverse problems in science and engineering. Springer, URL <http://arxiv.org/abs/0912.3522v4>.

[39] V. Thirumalai, P. Frossard, Image reconstruction from compressed linear measurements with side information, in: Proceedings of IEEE International Conference on Image Processing, 2011.