JOINT RECONSTRUCTION OF CORRELATED IMAGES FROM COMPRESSED LINEAR MEASUREMENTS

Vijayaraghavan Thirumalai and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory - LTS4 , Lausanne, 1015 - Switzerland.
Email:{vijayaraghavan.thirumalai, pascal.frossard}@epfl.ch

ABSTRACT

This paper proposes a joint reconstruction algorithm for compressed correlated images that are given under the form of linear measurements. We first propose a geometry based model in order to describe the correlation between visual information in a pair of images, which is mostly driven by the translational motion of objects or vision sensors. We consider the particular problem where one image is selected as the reference image and it is used as the side information for decoding the compressed correlated images. These compressed images are built on random measurements that are further quantized and entropy coded. The joint decoder first captures the most prominent visual features in the reference image using geometric basis functions. Since images are correlated, these features are likely to be present in the compressed images too, possibly with some small transformation. Hence, the reconstruction of the compressed image is based on a regularized optimization problem that estimates these features in the compressed images. The regularization term further enforces the consistency between the reconstructed images and the quantized measurements. Experimental results show that the proposed scheme is able to efficiently estimate the correlation between images. It further leads to good reconstruction performance. The proposed scheme is finally shown to outperform DSC schemes based on unsupervised disparity or motion learning as well as independent coding solution based on JPEG-2000 from a rate-distortion perspective.

1. INTRODUCTION

Distributed source coding (DSC) usually refers to the independent encoding and joint decoding of correlated sources. It permits to design low complexity acquisition systems and to shift the computational burden to the decoder. DSC typically finds applications in vision sensor networks where low-power cameras perform a spatio-temporal sampling of the visual information and send the resulting images to a central decoder. While most common encoders in DSC systems acquire the entire image before compression, the complexity of the encoders can be further reduced if the sensors directly acquires the compressed image in the form of random projections [1, 2]. Such a solution computes only few linear projections at the encoder and thereby significantly reduces the computational cost and the power requirements at the encoder. A joint decoder eventually reconstructs the visual information from the compressed images by exploiting the correlation between the samples, which permits to achieve a good rate-distortion tradeoff in the representation of video or multi-view information.

Duarte *et al* [3] have proposed distributed compression of correlated signals from linear measurements. In particular, three joint sparsity models are proposed to exploit the correlation between signals at decoder and are used in joint signal reconstruction algorithms. These simple joint sparsity models are however not ideal in the case of natural images. Later the concept of random projections has been then applied for distributed video coding in efforts to reduce the complexity of the encoding stage [4, 5, 6]. However, these coding schemes generally assume that the signal is sparse in a particular orthonormal basis (e.g., DCT or Wavelet) [4, 5] or in a block based dictionary [6]. It is more generic to assume the signal to be sparse in a structured redundant dictionary since this leads to greater flexibility in the choice of the representation of the signal and in the construction of the correlation model. Rauhut *et al* [7] extend the concept of signal reconstruction from linear measurements using redundant dictionaries, but this idea has not been extended to distributed scenarios.

In [8], we studied the problem of estimating the correlation model between a reference image and a highly compressed image, where the visual information for the compressed image is given in the form of few quantized linear measurements. In this paper, we build on our previous work [8] and propose a joint reconstruction algorithm, which estimates the correlation model as well as, reconstructs the highly compressed image using the estimated correlation model. We first compute the most prominent visual features in the reference image and approximate them with geometric functions drawn from a parametric dictionary. Since the images are correlated, the geometric features are likely to appear in compressed images, possibly after some simple transformations. We then formulate a regularized optimization framework whose objective is to compute the visual features in the compressed image, under the assumption that they represent shifted versions of visual features in the reference image. We add a regularization constraint in order to ensure the reconstructed compressed image to be consistent with the quantized measurements. At the same time we also enforce the consistency of the motion information contained by our correlation model. We show by experiments that the proposed algorithm computes a good estimation of the motion or disparity field between the pair of images in video or multiview scenarios, respectively. We also show that the inclusion of the consistent reconstruction term in the optimization model is very effective in improving the reconstruction quality of the compressed image. In particular, we show that the rate-distortion (RD) performance of the proposed scheme outperforms DSC scheme based on unsupervised disparity or motion learning [9] and independent coding scheme like JPEG 2000. Finally, we show the benefit of geometry based structured dictionaries compared to adaptive dictionary built on patches from the reference image [6] for the joint reconstruction of correlated image pairs.

2. PROPOSED FRAMEWORK

We consider a framework where a pair of images I_1 and I_2 that represent a scene at different time instants or from different viewpoints. The images are correlated through the motion of visual objects. They are transmitted to a joint decoder that estimates the relative motion or disparity between the received signals for efficient joint reconstruction. The framework is illustrated in Fig. 1.

One of the images is encoded and decoded independently and serves as a reference image for the joint reconstruction. While this image could be encoded with any coding algorithm, we choose here to represent the reference image I_1 by random linear measurements $y_1 = \psi I_1$ with a projection matrix ψ. The measurements are used by the decoder to reconstruct an approximation $\hat{I}_1$ using a convex optimization algorithm [10] under the assumption that I_1 is sparse in particular basis (e.g., a Wavelet basis). The second image I_2 is

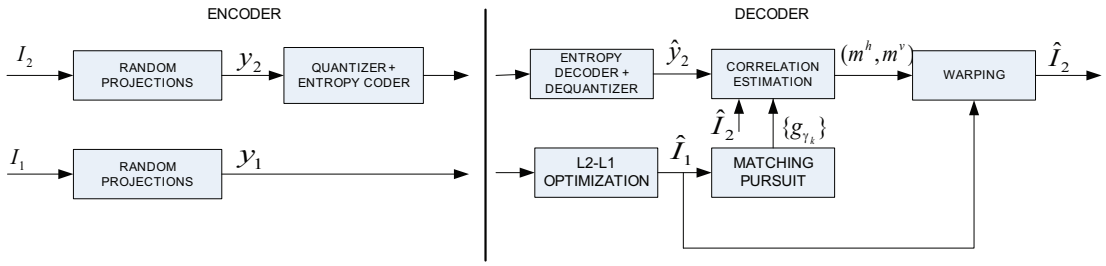


Figure 1: Schematic representation of the proposed scheme. The images I_1 and I_2 are correlated through displacement of scene objects, due to view point change or motion of scene objects.

also projected on a random matrix ψ to generate the measurements $y_2 = \psi I_2$. The generated measurements y_2 are further quantized using an uniform quantizer and are further entropy coded (e.g., Arithmetic encoder). The decoder performs the reverse operations (dequantization and entropy decoding) to form the measurement vector $\hat{y}_2$ (see Fig. 1). This measurement vector is finally used by the joint decoder to estimate the relative transformation between the images I_1 and I_2 and eventually reconstruct the second image $\hat{I}_2$.

We propose to model the correlation between the images by relative transformation between prominent visual features in both images. We assume that the images I_1 and I_2 can be represented by a sparse linear expansion of geometric function g_γ taken from a parametric and overcomplete dictionary $D = \{g_\gamma\}$. The geometric function g_γ in D is usually called an *atom*. The dictionary is constructed by applying a set of geometric transformations to the generating function g. These geometric transformations can be represented by a family of unitary operator $U(\gamma)$, so that the dictionary spanning the input space takes the form $D = \{g_\gamma = U(\gamma)g, \gamma \in \Gamma\}$ for a given set of transformation indexes Γ. Typically this transformation set consists of scaling s_x, s_y, rotation θ, and translation t_x, t_y operators, defined as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1/s_x & 0 \\ 0 & 1/s_y \end{bmatrix} \begin{bmatrix} cos\ \theta & sin\ \theta \\ -sin\ \theta & cos\ \theta \end{bmatrix} \begin{bmatrix} x - t_x \\ y - t_y \end{bmatrix} \quad (1)$$

where (x,y) defines the image coordinates. Thus, each of the transformation is indexed by five parameters.

We can then write the approximation of the decoded reference image $\hat{I}_1$ with functions in D as

$$\hat{I}_1 \approx \sum_{k=1}^{N} c_k\, g_{\gamma_k}. \quad (2)$$

where $\{c_k\}$ are the set of N coefficients. The approximation of $\hat{I}_1$ can be computed by sparse algorithms such as Matching Pursuit [11], which greedily picks up the N atoms $\{g_{\gamma_k}\}$ that best match the image $\hat{I}_1$. Under the assumption that the images I_1 and I_2 are correlated, the second image I_2 can be approximated with transformed versions of the atoms used in the approximation of $\hat{I}_1$. We can thus write

$$I_2 \approx \sum_{k=1}^{N} c_k\, F^k(g_{\gamma_k}), \quad (3)$$

where $F^k(g_{\gamma_k})$ represents a local geometrical transformation applied to the atom g_{γ_k}. Due to the parametric form of the dictionary, the effect of F^k corresponds to a geometrical transformation of the atom g_{γ_k} that results in another atom in the same dictionary D. Therefore, it is interesting to note that the transformation F^k on g_{γ_k}, boils down to a transformation of the atom parameters, i.e.,

$$F^k(g_{\gamma_k}) = U(\delta\gamma)g_{\gamma_k} = U(\gamma_k \circ \delta\gamma)g = g_{\gamma_k \circ \delta\gamma} = g_{\gamma'_k}. \quad (4)$$

Now, the main challenge in the joint decoder is to estimate the local geometrical transformation F^k for each of the atom g_{γ_k} in $\hat{I}_1$ from the linear measurements $\hat{y}_2$. We formulate in the next section a regularized optimization problem in order to estimate F^k, or equivalently the relative motion or disparity between images I_1 and I_2 that leads to an efficient representation of the image $\hat{I}_2$.

3. JOINT RECONSTRUCTION FROM COMPRESSED LINEAR MEASUREMENTS

Given the set of N atoms $\{g_{\gamma_k}\}$ that approximate the first image $\hat{I}_1$ the joint reconstruction problem consists first in finding the corresponding visual patterns in the second image I_2, while the later is given only by compressed random measurements $\hat{y}_2$. This is equivalent to finding the correlation between the images with the joint sparsity model described in Eq. 3. This correlation is eventually used to reconstruct the compressed image.

3.1 Regularized Energy Model

The main challenge is to estimate the set of N atoms in the second image I_2 that correspond to the set of visual features in the reference images given by their atom parameters $\{\gamma_k\}$. For our convenience we denote the set of N atom parameters in I_2 by Λ, where $\Lambda = (\gamma'_1, \gamma'_2, ...\gamma'_N)$. We propose to estimate this set of parameters in a regularized energy minimization framework. The energy model E proposed in our scheme is expressed as

$$E(\Lambda) = E_d(\Lambda) + \alpha_1 E_s(\Lambda) + \alpha_2 E_t(\Lambda), \quad (5)$$

where E_d, E_s and E_t represent the data term, smoothness term and reconstruction term respectively. The regularization constants α_1 and α_2 balance the data, smoothness and reconstruction terms. The solution to the correlation estimation (for efficient reconstruction of $\hat{I}_2$) is given by the set of N atom parameters Λ^* that minimizes the energy E, i.e.,

$$\Lambda^* = arg\min_{\Lambda \in S} E(\Lambda) \quad (6)$$

where S represents the search space. The search space S is given by

$$S = \{(\gamma'_1, \gamma'_2, ...\gamma'_N) \mid \gamma'_k = \gamma_k + \delta\gamma, 1 \le k \le N, \delta\gamma \in \mathscr{U}\}. \quad (7)$$

where $\mathscr{U} \subset \mathbb{R}^5$, and $\mathscr{U} = [-\delta t_x\ \delta t_x] \times [-\delta t_y\ \delta t_y] \times [-\delta\theta_x\ \delta\theta_x] \times [-\delta s_x\ \delta s_x] \times [-\delta s_y\ \delta s_y]$ where $\delta t_x, \delta t_y, \delta\theta_x, \delta s_x, \delta s_y$ are the search window sizes corresponding to translation parameters t_x, t_y, rotation θ and scales s_x, s_y respectively.

Now we turn our attention in describing the three cost functions used in Eq. 5. Given the set of N atom parameters $\Lambda = \{\gamma'_k\}$, the data cost function E_d measures the error between the quantized measurements $\hat{y}_2$ and the orthogonal projection of $\hat{y}_2$ onto the columns spanned by Ψ_Λ, where $\Psi_\Lambda = \psi[g_{\gamma'_1}|g_{\gamma'_2}|.....|g_{\gamma'_N}]$. It turns out that the orthogonal projection operator $\mathscr{P}$ is given by $\mathscr{P} = \Psi_\Lambda \Psi_\Lambda^\dagger$, where $\Psi_\Lambda^\dagger$ represents the pseudo-inverse. Therefore the data term estimates the set of N atom parameters Λ that minimizes the mean

square error (MSE) w.r.t. quantized measurements $\hat{y}_2$. More formally, the data cost E_d is computed as

$$E_d(\Lambda) = \| \hat{y}_2 - \Psi_\Lambda \Psi_\Lambda^\dagger \hat{y}_2 \|_2. \quad (8)$$

Before describing the smoothness term E_s, we discuss here the estimation of dense motion field from the atom transformation. Given a pair of corresponding atoms g_{γ_k} and $g_{\gamma'_k}$ in the images I_1 and I_2 respectively, we first calculate the mapping of each pixel $\mathbf{z} = (x,y)$ in g_{γ_k} to its corresponding pixel $\tilde{\mathbf{z}} = (\tilde{x},\tilde{y})$ on $g_{\gamma'_k}$ using Eq. 1. This grid transformation $\mathbf{z}^{(k)} - \tilde{\mathbf{z}}^{(k)} = (x^{(k)} - \tilde{x}^{(k)}, y^{(k)} - \tilde{y}^{(k)})$ corresponds to the amount of local motion captured by the k^{th} pair of atoms g_{γ_k} and $g_{\gamma'_k}$. Using a similar process, the mapping is established for all atom pairs from the respective transform parameters γ_k and γ'_k. Then the grid transformation captured by all the N pairs of atom are fused together to estimate the dense motion field. In the fusion process, we simply take the most confident transformation or motion $\mathbf{z}^{(k)} - \tilde{\mathbf{z}}^{(k)}$ for each location $\mathbf{z}$, from the set of transformations $\{\mathbf{z}^{(k)} - \tilde{\mathbf{z}}^{(k)}\}$ induced by the N atoms. We first assign weights $\{w_{\mathbf{z}}^{(k)}\}$ based on the response of the k^{th} atom at the pixel location $\mathbf{z}$. Then the most confident mapping or equivalently the motion at location $\mathbf{z}$ is assigned as $\mathbf{z}^{(\bar{k})} - \tilde{\mathbf{z}}^{(\bar{k})}$, where $\bar{k}$ $(1 \leq \bar{k} \leq N)$ is the atom index for which the maximum weight is found out. Thus the horizontal and vertical components of the motion field at location $\mathbf{z}$ is given by

$$(\mathbf{m}^h(\mathbf{z}), \mathbf{m}^v(\mathbf{z})) = (x^{(\bar{k})} - \tilde{x}^{(\bar{k})}, y^{(\bar{k})} - \tilde{y}^{(\bar{k})}) \quad (9)$$

where $\bar{k} = arg \max_{k=1,2,...N} w_{\mathbf{z}}^{(k)}$, and $w_{\mathbf{z}}^{(k)}$ is the response of the k^{th} atom at the location $\mathbf{z}$. i.e., $w_{\mathbf{z}}^{(k)} = g_{\gamma_k}(\mathbf{z}) = g_{\gamma_k}(x,y)$.

We can now compute the smoothness term E_s, whose objective is to create a consistent correlation estimation between images. We generate a dense motion (or disparity) field from the atom transformation, and later we penalize the motion (or disparity) field to be coherent among adjacent pixels. We compute the smoothness cost E_s using,

$$E_s = \sum_{\mathbf{z},\mathbf{z}' \in \mathcal{N}} V_{\mathbf{z},\mathbf{z}'} \quad (10)$$

where $\mathbf{z},\mathbf{z}'$ are the adjacent pixel locations and $\mathcal{N}$ is the usual 2 pixel neighborhood. The term $V_{\mathbf{z},\mathbf{z}'}$ in Eq. 10 is defined as,

$$V_{\mathbf{z},\mathbf{z}'} = min(|\mathbf{m}^h(\mathbf{z}) - \mathbf{m}^h(\mathbf{z}')| + |\mathbf{m}^v(\mathbf{z}) - \mathbf{m}^v(\mathbf{z}')|, K) \quad (11)$$

where $\mathbf{m}^h(\mathbf{z})$, and $\mathbf{m}^v(\mathbf{z})$ represent the horizontal and vertical components of the motion field respectively at the pixel location $\mathbf{z} = (x,y)$, and the parameter K is a constant. The parameter K sets a maximum limit to the penalty, and thus helps to preserve the discontinuities in the motion field [12].

Finally, we further improve the reconstruction quality of the image $\hat{I}_2$ by adding a reconstruction term E_t to the energy model described in Eq. 5. The term E_t calculates the l_2 norm error between the measurements generated from the reconstructed image $\hat{I}_2$ and quantized measurements $\hat{y}_2$. In other words, the cost function E_t enforces the reconstructed image $\hat{I}_2$ to be consistent with the quantized measurements $\hat{y}_2$. The reconstruction term E_t is computed as

$$E_t = \| \hat{y}_2 - \mathcal{Q}[\psi \mathcal{W}(\hat{I}_1)] \| = \| \hat{y}_2 - \mathcal{Q}[\psi \hat{I}_2] \| \quad (12)$$

where $\mathcal{Q}$ is the quantizer and $\mathcal{W}$ warps the reference image $\hat{I}_1$ using the generated motion or disparity field (see Fig. 1).

Finally, note that in general the transformation F^k acting on atom g_{γ_k} might change the position (t_x,t_y), rotation θ and scales s_x,s_y of the atom g_{γ_k} or could be any one or combination of these changes. In this work, we approximate the transformation F^k to act only on the integer locations of the translational component (t_x,t_y) of the atom g_{γ_k} as our correlation model is based on atom shift that approximate the motion of objects in the scene. We experimentally show in the next section that such an approximation in the transformation F^k gives a good estimation of the correlation model between the images.

3.2 Optimization algorithm

We describe here the optimization methodology to solve Eq. 6 and estimate the transformation F_k for each of the atom. One trivial approach would be to perform an exhaustive search on the entire search space S to estimate the solution. But the cost of such a solution is high, as the number of elements in the search space S grows exponentially with the window size $\delta t_x, \delta t_y$ i.e., $|\tilde{S}| = N^{((2\delta t_x+1)\times(2\delta t_y+1))}$. We rather propose a suboptimal solution that estimates the transformations F_k iteratively, by deforming each of the N atom parameters γ_k by one increment in the parameter space. In particular, as we search for translational motion, we focus on the search space that is given by perturbing the translational components t_x and t_y of each atom position by one unit i.e., $t_x \pm 1$, $t_y \pm 1$ for each atom γ_k. We first initialize the algorithm with zero motion, i.e., the atoms $\{g_{\gamma_k}\}$ generated from $\hat{I}_1$ are used in the first iteration, $\{\gamma'_k\} = \{\gamma_k\}$, and the search space is S' is formed using

$$\begin{aligned} S' = \{&(\gamma'_1, \gamma'_2, ..., \hat{\gamma}'_k, ..., \gamma'_N) | \hat{\gamma}'_k = (t_x^k + j_1, t_y^k + j_2, \theta^k, s_x^k, s_y^k), \\ &1 \leq k \leq N, j_1, j_2 \in \mathbb{Z}, -1 \leq j_1, j_2 \leq 1\} \subset S. \end{aligned} \quad (13)$$

We then calculate the energy E in Eq. 5 for the set of N atoms in the search space S'. It can be easily shown that the size of the search space S' is at most $8N+1$, i.e.,$|S'| = 8N+1$. Once the energy E is computed for atoms in S', we find the parameters $(\gamma'_1, \gamma'_2, ..., \gamma'_N)$ corresponding to the minimum energy. Then a new search space S' is formed using Eq. 13 with the current parameter solution $(\gamma'_1, \gamma'_2, ..., \gamma'_N)$, and this procedure is repeated until convergence is reached. The joint decoding algorithm is summarized in Algorithm 1.

The proposed algorithm is guaranteed to converge. Let E_0 be the initial energy i.e., the energy corresponding to set of parameters $\gamma'_k = \gamma_k, \forall k$ where $1 \leq k \leq N$. As described above, in the first iteration we form the search space S' using Eq. 13 and then the set of atom parameters corresponding to the minimum energy is computed. Let E_1 be the corresponding minimum value of the energy found in the first iteration. It is clear that $E_1 \leq E_0$, as the search space S' includes the initial set of parameters $\gamma'_k = \gamma_k, \forall k$ where $1 \leq k \leq N$. By using the same argument, we conclude that $E_i \leq E_{i-1}$, where E_i and E_{i-1} are the minimum value of the energy corresponding to the iteration i, and $i-1$ respectively. As $E_i < E_{i-1}$, we therefore conclude that the energy continues to decrease for every iteration till it reaches a local or global minima E_{min}. When $E_i = E_{min}$ for some iteration number i, the energy cannot decrease beyond E_{min}, and therefore it remains constant i.e., $E_i = E_{i+1} = E_{min}$. Thus we conclude that the proposed optimization scheme converges to a local or global minima and allows us to estimate a suboptimal solution with tractable computational complexity.

4. EXPERIMENTAL RESULTS

The scheme we proposed is generic and it can be applied for estimating the disparity from two cameras or the motion field from two frames in a video sequence. In this section, we present the experimental results for both applications.

4.1 Disparity Estimation from stereo cameras

We evaluate the performance of our scheme using Sawtooth image set [1] with a resolution 144×176 pixels. As the images are rectified,

[1]These image sets are available in http://vision.middlebury.edu/stereo/data/. The image sets are then

Algorithm 1 Joint Decoder

1: Input N, α_1, α_2, K, δt_x, δt_y
2: Generate $\{g_{\gamma_k}\}$ from $\hat{I}_1$ s.t. $\hat{I}_1 \approx \sum_{k=1}^{N} c_k\, g_{\gamma_k}$
3: Initialize $(\mathbf{m}^h, \mathbf{m}^v) = (\mathbf{0}, \mathbf{0})$ i.e., $\{\gamma'_k\} = \{\gamma_k\}$
4: *repeat*
5: Generate index search space S' using Eq. 13
6: **for** 1: $|S'|$ **do**
7: Calculate the energy E
8: **end for**
9: Estimate the N atoms indexes $\{\gamma'_k\}$ corresponding to the minimum energy.
10: *until convergence is reached*

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

Figure 2: (a) Disparity field $\mathbf{m}^h$ generated in our scheme from 8870 quantized measurements (corresponds to 35% measurement rate) (b) Error in the disparity field (white pixels denote error).

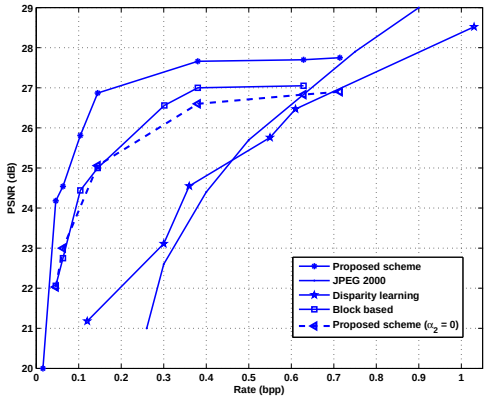


Figure 3: RD comparison of the proposed scheme w.r.t. DSC scheme [9], block based scheme [6] and independent coding solution based on JPEG 2000.

the disparity estimation problem is simplified to a one dimensional search problem. Therefore, the disparity map can be represented by the horizontal component $\mathbf{m}^h$ of the field given in Eq. 9, with no changes observed in the vertical component $\mathbf{m}^v$ i.e., $\mathbf{m}^v = \mathbf{0}$. In our scheme, the dictionary is constructed using two generating functions, as explained in [11]. The first one consists of 2D Gaussian functions, to capture low frequency component. The second function represents Gaussian in one direction, and the second derivative of 2D gaussian in the orthogonal direction to the capture edges. The translation parameters t_x and t_y take any positive value from one to the resolution of the image i.e., t_x varies from 1 to 176, while t_y varies from 1 to 144. Ten rotation parameters are used between 0 and π, with increments $\pi/18$. Five scaling parameters are equidistributed in the logarithmic scale from 1 to $N_1/8$ vertically, and 1 to $N_2/9.77$ horizontally, where $N_1 \times N_2$ is the size of the image.

The random projections are computed using the scrambled Hadamard ensemble with block size of 8 [10]. The measurements y_2 are quantized uniformly using a two bit quantizer and further encoded using an arithmetic coder. The rate control at the encoder is achieved by varying the measurement rate or the number of projections y_2. The reference view I_1 is encoded such that the quality of $\hat{I}_1$ is approximately 33 dB. Matching Pursuit is carried out on $\hat{I}_1$, and the image $\hat{I}_1$ is approximated using $N = 60$ atoms. In our experiments, the number of atoms N is chosen in such a way, that the selected N prominent features covers the entire scene given by the image I_1. The search for the transformation F^k is carried out along the translational component t_x with window size $\delta t_x = 4$ pixels, and no changes are considered along the translational component t_y.

For a given measurement rate, we first estimate the disparity field using the procedure described in Algorithm 1. Fig. 2(a) shows the estimated disparity field $\mathbf{m}^h$ from 8870 quantized measurements (corresponds to 35% measurement rate). We then compare our results w.r.t. ground truth $\mathbf{M}^h$ (available in http://vision.middlebury.edu/stereo/data/) and the comparison is available in Fig. 2(b). From Fig. 2(b) it is clear that the proposed scheme gives a good estimation of the disparity field where the error is localized along the edges due to the choice of the dictionary function. Then the estimated disparity field is used to reconstruct the image $\hat{I}_2$ by warping the reference image $\hat{I}_1$. Fig. 3 shows the RD comparison for the reconstructed image $\hat{I}_2$ w.r.t. JPEG 2000 based coding strategy. It is clear that our scheme outperforms independent coding solution based on JPEG 2000 at low to medium rates, due to the efficient joint reconstruction. However, our coding performance saturates at 0.7 bpp, as the fine details or texture in the scene cannot be captured by the proposed scheme due to the choice of dictionary functions and the limits of the correlation model in capturing non-structural components. We then compare the RD performance of the reconstructed image $\hat{I}_2$ with, and without activating the reconstruction term E_t (corresponds to $\alpha_2 = 0$ in Eq. 5), and the comparison is available in Fig. 3. From Fig. 3 it is clear that the quality of the image $\hat{I}_2$ is improved by enabling the reconstruction term E_t.

We then compare this performance to a DSC scheme, where the disparity field is estimated at the decoder using Expected Maximization (EM) principles [9]. To have a fair comparison, we encode the reference image using similar principles described in section 2 and the quality of the image $\hat{I}_1$ in the joint decoder is 33 dB. The image I_2 is first transformed using 8×8 DCT, and the resulting coefficients are quantized. The quantized DCT coefficients are further encoded using LDPC channel codes, and the resulting syndromes are transmitted to the joint decoder. The joint decoder uses $\hat{I}_1$ as the side information, and estimates the disparity from the syndromes using an unsupervised learning scheme via EM. Finally the image $\hat{I}_2$ is reconstructed by compensating the disparity in the reference image $\hat{I}_1$. Fig. 3 compares the quality of the reconstructed image $\hat{I}_2$ with our scheme. From Fig. 3 it is clear that the proposed scheme outperforms the DSC coding scheme based on EM principles.

Finally, in order to demonstrate the benefit of geometric dictionary, we compare the results to a scheme that adaptively constructs the dictionary using blocks or patches in the reference image [6]. In our experiments, we construct a dictionary in the joint decoder from the reference image $\hat{I}_1$ using 8×8 blocks. We then used the optimization scheme described in algorithm 1 to select the best block from the adaptive dictionary, with a search window size of $\delta t_x = 4$ pixels along the horizontal direction. Fig. 3 shows the quality of reconstruction for such a solution, and it is clear that our scheme outperforms block-based dictionaries mainly due to rich representation of the visual information provided by the structured dictionary.

4.2 Motion estimation from video sequence

We further study the performance of our scheme for the motion estimation problem in video sequences. We built the image set using

downsampled to a resolution 144×176 using bilinear filters.

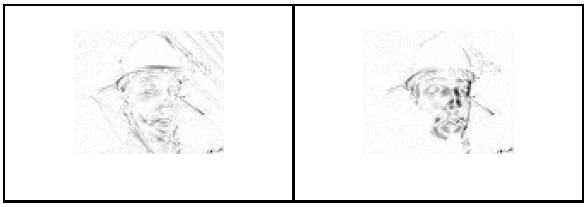
the frames 2 and 3 of the Foreman sequence. The frame 2 is selected as the reference image I_1, and approximated with a quality of approx. 45 dB in the joint decoder. We used the dictionary described in the previous section for approximating the image $\hat{I}_1$. For this particular data set, we approximate $\hat{I}_1$ using $N = 60$ atoms. The search window size is $\delta t_x = \delta t_y = 4$ pixels for both the translational components t_x and t_y.

Fig. 4(a) and Fig. 4(b) compare the residual energy of the reconstructed image $\hat{I}_2$ w.r.t. I_2 and I_1 respectively. The MSE between the images $\hat{I}_2$ and I_1 is 73, while the MSE between $\hat{I}_2$ and I_2 is 41, and this indicates that the proposed scheme efficiently captures the correlation between I_1 and I_2. The RD performance of the reconstructed image $\hat{I}_2$ is shown in Fig. 5, and it is then compared to JPEG 2000, DSC and block-based schemes. From the plot is clear that our coding scheme outperforms these competitors due to efficient joint reconstruction. Also from Fig. 5 we observe that the quality of $\hat{I}_2$ is improved by activating the reconstruction term E_t ($\alpha_2 \neq 0$). It should be noted that when $\alpha_2 \neq 0$ in Eq. 5, we estimate only the motion field (i.e., no joint reconstruction) as described in our previous work [8].

Finally, we compare our results with a joint encoding scheme based on H.264, with GOP size 2. In H.264 scheme, the image I_1 is selected as the reference frame, and it is approximated to 45 dB in the joint decoder. We then vary the quantization parameter for the frame I_2, and the image $\hat{I}_2$ is reconstructed. We carry out this experiment in two different settings, (1) variable macro block size (H.264 - variable block size) (2) fixed macro block size 8×8 (H.264 - block size 8). The corresponding RD plot for the two cases are available in Fig. 5. From Fig. 5, we could infer that our scheme performs better than the H.264 scheme especially at low rates, when a fixed macro block size is used for motion estimation. As the proposed scheme fails to capture the fine details or the texture, we are 4 dB (approx) far from the H.264 scheme at higher rates.

5. CONCLUSIONS

In this paper we have presented a methodology to compute the joint reconstruction of the compressed image pairs from quantized linear measurements. We have used a geometry based structured dictionary to capture the prominent geometric features in the images. We have related the corresponding features in the images using a geometry based correlation model under translational motion assumptions. Experimental results demonstrate that the proposed methodology computes a good estimation of dense disparity or motion field. We have also demonstrated that the geometry based dictionary captures effectively the correlation between frames, comparing to an adaptive block based dictionary. We have also shown that the regularization term based on consistent reconstruction is quite efficient in improving the quality of the reconstructed image. Finally, the proposed scheme outperforms JPEG 2000 and DSC schemes in terms of RD performance, which positions it as an effective solution for distributed image processing with low encoding complexity.



(a) MSE: 41 **(b) MSE: 73**

Figure 4: Comparison of reconstructed image $\hat{I}_2$ w.r.t. I_2 and I_1 (a) $1-|\hat{I}_2 - I_2|$ (b) $1-|\hat{I}_2 - I_1|$ (white pixel denotes no error). The image $\hat{I}_2$ is reconstructed using 3801 quantized measurements.

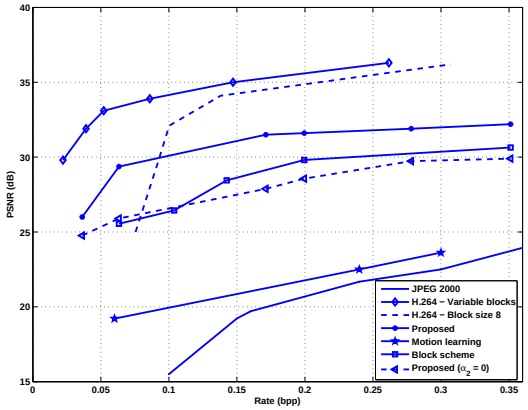


Figure 5: RD comparison of the proposed scheme w.r.t. block based scheme [6], JPEG 2000, DSC [9] and H.264 coding schemes.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Infor. Theo.*, vol. 52, pp. 1289–1306, 2006.

[2] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Infor. Theo.*, vol. 52, pp. 489–509, 2006.

[3] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conf. on Sig. Sys. and Comp.*, 2005.

[4] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *Proc. IEEE ICASSP*, 2009.

[5] T. T. Do, Y. Chen, D. T. Nguyen, N. Nguyen, L. Gan, and T. D. Tran, "Distributed compressed video sensing," in *Proc. IEEE ICIP*, 2009.

[6] J. P. Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. PCS*, 2009.

[7] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Infor. Theo.*, vol. 54, pp. 2210–2219, 2006.

[8] V. Thirumalai and P. Frossard, "Motion estimation from compressed linear measurements," in *Proc. IEEE ICASSP*, 2010.

[9] D. Varodayan, D. Chen, M. Flierl, and B. Girod, "Wyner-ziv coding of video with unsupervised motion vector learning," *EURASIP Signal Processing: Image Communication*, vol. 23, pp. 369–378, 2008.

[10] L. Gan, T. T. Do, and T. D. Tran, "Fast compressive imaging using scrambled hadamard ensemble," in *Proc. EUSIPCO*, 2008.

[11] R. M. Figueras, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. Image Proc.*, vol. 15, pp. 726–739, 2006.

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