

DENSE DISPARITY ESTIMATION FROM LINEAR MEASUREMENTS

Vijayaraghavan Thirumalai and Pascal Frossard

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

ABSTRACT

This paper proposes a methodology to estimate the correlation model between a pair of images that are given under the form of linear measurements. We consider an image pair whose common objects are relatively displaced due to the positioning of vision sensors. In such scenarios the correlation model that relates the displacement between the objects is effectively represented by a disparity image. We consider a framework where each image is directly acquired and compressed by projecting onto a random basis of lower dimension. Given the linear measurements computed from the images we propose to estimate the underlying correlation model directly in the compressed domain without reconstructing the images that is usually a costly solution. We first show that the correlated images can be efficiently related using a linear operator. Using this linear relationship between the images we derive the relationship between the corresponding measurements in the compressed domain. The underlying correlation model is then built by solving a regularized energy minimization problem. Experimental results show that the proposed scheme estimates an accurate correlation model between the images. Also we show by experiments that the proposed scheme performs competitively with the scheme that estimates the correlation model from the reconstructed images.

1. INTRODUCTION

Distributed processing has recently found numerous applications in vision sensor networks. In practice the vision sensors deployed in such networks acquire the entire image before compression. With the advent of compressed sensing it becomes possible to directly acquire the compressed image in the form of random projections [1, 2]. Such a scheme computes only few linear projections at the encoder and thereby significantly reduces the computational cost and the power requirements at the encoder. Given the visual information in terms of linear measurements, one of the most important and challenging tasks in distributed processing is to estimate the correlation between the signals or images captured by different sensors, so that the information can be efficiently processed, coded or rendered. Furthermore in applications like detection or rendering it is sufficient to estimate only the correlation model and the explicit reconstruction of images is not required. This motivates us to estimate the correlation model directly from the linear measurements without reconstructing the references images.

The concept of random projections in distributed scenarios has been previously studied in [3] where three joint sparsity models are designed and used in joint signal reconstruction algorithms. 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. Later the concept of random projections has been applied to distributed video coding [4, 5] or multiview coding [6] in efforts to reduce the complexity of the encoding stage. However these schemes estimate correlation models from the reconstructed images and not directly from the linear measurements. In other words, the reference images are first reconstructed independently from linear measurements by solving an $l_2 - l_1$ optimization problem, and then a correlation model is estimated from the reconstructed images. In our previous work [7] we have also proposed a methodology to estimate the correlation model from a reference image and highly compressed linear measurements. Unfortunately, reconstructing the reference images based on an $l_2 - l_1$ optimization problem is highly complex. Hence we propose to estimate the correlation model directly from the linear measurements without any intermediate reconstruction stage.

In this paper we consider an image pair captured by a stereo camera. In such a scenario the common objects between the images are displaced due to change in view point and the correlation model is usually given in the form of a dense disparity image. We propose to estimate this correlation model directly from the linear measurements in a regularized energy minimization framework. The data cost function assigns a particular disparity value to a pixel which best agrees with the linear measurements. Then a smoothness cost function penalizes the disparity value for the adjacent pixels so that they stay as close as possible, except at discontinuities. Experimental results show that the proposed scheme is able to efficiently estimate the correlation model between the images. Also the performance of the proposed scheme closely matches the performance of the scheme that estimates a disparity from reconstructed images where the images are reconstructed from the linear measurements by solving an $l_2 - l_1$ optimization problem. Thus the proposed scheme provides an interesting solution for distributed processing in vision sensors which estimates the correlation model directly in the compressed domain from low complexity linear measurements.

2. PROPOSED FRAMEWORK

We consider a pair of image I_1 and I_2 that represent the same scene taken at different viewpoints, with resolution $N_1 \times N_2$. Without loss of generality, we assume that the images I_1 and I_2 are rectified so that the disparity estimation problem reduces to an one dimensional search problem. As the images are rectified we propose to compute the linear measurements across rows in order to facilitate the matching process, i.e., we consider each row as one block (with size $1 \times N_2$), and the measurements are independently taken for each block. Such block-based sampling scheme is commonly used in the literature for easy handling and sampling speed-up [8]. In more de-

This work has been partly supported by the Swiss National Science Foundation, under grant 200021-118230.

tails, let $I_{1,k}$ and $I_{2,k}$ represent the k^{th} row of 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 a measurement matrices ϕ_1^k and ϕ_2^k respectively. The measurements $Y_{1,k}$ and $Y_{2,k}$ are computed as,

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

where $(\cdot)^T$ denotes the transpose operator. It should be noted that ϕ_1^k and ϕ_2^k are of dimensions $M \times N_2$ with $M \ll N_2$.

Now we describe how to relate the images I_1 and I_2 using a linear operator. Let $D = [D_1, D_2, \dots, D_{N_1}]^T$ be a disparity image that describes the underlying correlation between the images, where D_k represents the k^{th} row of D . As the images are correlated through the displacement of scene objects it is likely that each pixel $\mathbf{z} = (x, y)$ in the first image might have shifted to $\mathbf{z}' = (x + D(x, y), y)$ in the second image with the disparity $D(x, y)$. Then the images I_1 and I_2 can be simply related by a linear operator T_D which changes the coordinate system from (x, y) to $(x + D(x, y), y)$, i.e.,

$$\begin{aligned} I_2(x, y) &= T_D\{I_1(x, y)\} \\ I_2(x, y) &= I_1(x + D(x, y), y). \end{aligned} \quad (2)$$

Eq. (2) can be extended to describe the relationship between the k^{th} row of the images $I_{1,k}$ and $I_{2,k}$. It is easy to check that the rows $I_{1,k}$ and $I_{2,k}$ can be related using D_k as,

$$I_{2,k} = I_{1,k+D_k}, \quad \forall k = 1, 2 \dots N_1. \quad (3)$$

For mathematical convenience we represent Eq. (3) as a linear system that relates the rows $I_{1,k}$ and $I_{2,k}$ using a matrix A^k . Mathematically it is given as,

$$I_{2,k}^T = A^k I_{1,k}^T, \quad \forall k = 1, 2 \dots N_1, \quad (4)$$

where the dimensions of A^k is $N_2 \times N_2$. The entries of the matrix A^k are given by

$$A^k(x, \min(x + \beta, N_2)) = \begin{cases} 1 & D_k(x) = \beta \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $D_k(x)$ represents the disparity value at the x^{th} location in the k^{th} row, i.e., the value at location $D(k, x)$. If the value of $x + \beta > N_2$ (which might happen at the boundaries) we replace $x + y = N_2$ so that the dimensions of the matrix A^k is $N_2 \times N_2$. It is easy to check that the matrix A^k formed using Eq. (5) contains only one 1 in each row. For example, the matrix A^k corresponding to $D_k = [2 \ 2 \ 1 \ 1]$ is given by

$$A^k = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6)$$

Since the matrix A^k contains only one 1 in each row it is evident from Eq. (4) that $I_{2,k}(i) = I_{1,k}(j)$ if $A^k(i, j) = 1$. Thus it is clear that the matrix A^k shifts the pixels in $I_{1,k}$ by its corresponding disparity vector D_k to form $I_{2,k}$. In the remaining section we focus on deriving the relationship between the measurements $Y_1 = [Y_{1,1}, Y_{1,2}, \dots, Y_{1,N_1}]^T$ and $Y_2 = [Y_{2,1}, Y_{2,2}, \dots, Y_{2,N_1}]^T$ when the images I_1 and I_2 are related linearly as given in Eq. (4).

Without loss of generality we assume that the measurements $Y_{1,k}$ and $Y_{2,k}$ are related by a linear operator B^k , i.e.,

$$Y_{2,k} = B^k Y_{1,k}, \quad \forall k = 1, 2 \dots N_1, \quad (7)$$

where the dimensions of B^k is $M \times M$. In fact one can show that any two vectors $Y_{1,k}, Y_{2,k} \in \mathbb{R}^M$ can be related using an operator B^k when $Y_{1,k} \neq \mathbf{0}$. However this linear operator B^k is a very generic one that is independent of the operator A^k that shifts the pixels between I_1^k and I_2^k . We thus derive the relationship between the matrixes B^k and A^k which is further used to derive the relationship between the measurements Y_1 and Y_2 . Pre-multiplying Eq. (1) by ϕ_2^k on both sides results in

$$Y_{2,k} = \phi_2^k I_{2,k}^T = \phi_2^k A^k I_{1,k}^T, \quad \forall k = 1, 2 \dots N_1. \quad (8)$$

Using Eq. (1), Eq. (7) can be rewritten as

$$Y_{2,k} = B^k Y_{1,k} = B^k \phi_1^k I_{1,k}^T, \quad \forall k = 1, 2 \dots N_1. \quad (9)$$

By equating Eq. (8) and Eq. (9) the matrixes A^k and B^k can be related as,

$$B^k \phi_1^k = \phi_2^k A^k, \quad \forall k = 1, 2 \dots N_1. \quad (10)$$

The above Eq. (10) forms an over-determined system of linear equations with M^2 unknowns (in matrix B^k) and $M \times N_2$ equations. In such scenarios it is well known that the best \hat{B}^k that minimizes $\|B^k \phi_1^k - \phi_2^k A^k\|_2$ is given by

$$\hat{B}^k = \phi_2^k A^k \phi_1^{k\dagger}, \quad \forall k = 1, 2 \dots N_1. \quad (11)$$

where \dagger denotes the pseudo-inverse operator. Using this result in Eq. (7) we get

$$Y_{2,k} \approx \phi_2^k A^k \phi_1^{k\dagger} Y_{1,k}, \quad \forall k = 1, 2 \dots N_1. \quad (12)$$

We thus show that if the k^{th} row in the images I_1 and I_2 are related linearly as given in Eq. (4) then the corresponding measurements $Y_{1,k}$ and $Y_{2,k}$ can be related linearly as given in Eq. (12). In the next section we describe the proposed methodology of estimating the disparity image D directly from the linear measurements.

3. DISPARITY ESTIMATION FROM LINEAR MEASUREMENTS

The objective of the dense disparity estimation is to assign a disparity value to each pixel $\mathbf{z} = (x, y)$ taken from a finite set $\mathcal{L} = \{l_1, l_2, \dots, l_{max}\}$. In other words our aim is to find a mapping f that assigns a disparity value $D(x, y) \in \mathcal{L}$ to each pixel \mathbf{z} such that the mapping f is piecewise smooth and also consistent with the given measurement vectors Y_1 and Y_2 . We propose to compute the optimal mapping f^* in an regularized energy minimization framework where the energy $E(f)$ is composed of a data term $E_d(f)$ and a smoothness term $E_s(f)$. By balancing the data and smoothness terms with $\lambda \geq 0$ the energy model can be represented as,

$$E(f) = E_d(f) + \lambda E_s(f). \quad (13)$$

The optimal mapping f^* can be found out by minimizing the energy function $E(f)$, i.e.,

$$f^* = \arg \min_f E(f). \quad (14)$$

The smoothness cost measures the penalty of assigning different labels or disparity to the adjacent pixels. It is computed using

$$E_s = \sum_{\mathbf{z}, \mathbf{z}' \in \mathcal{N}} V_{\mathbf{z}, \mathbf{z}'} \quad (15)$$

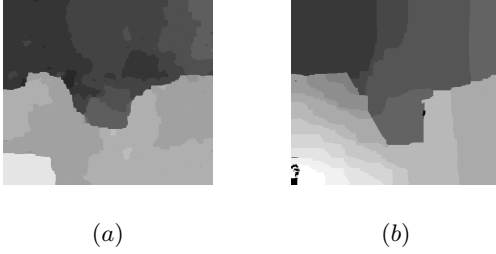


Fig. 1. Venus dataset: Computed disparity map D at a measurement rate (a) 0.2 (b) 0.7. The disparity error w.r.t. ground truth is found to be 41% and 10% respectively.

where \mathbf{z}, \mathbf{z}' are the pixels in the usual four pixel neighbourhood \mathcal{N} and the term $V_{\mathbf{z}, \mathbf{z}'}$ is given as,

$$V_{\mathbf{z}, \mathbf{z}'} = \min(|D(\mathbf{z}) - D(\mathbf{z}')|, K) \quad (16)$$

where K sets an upper level to the penalty that helps to preserve the discontinuities [9].

The data cost function measures the cost of assigning a particular label to the pixel based on the observations Y_1 and Y_2 . In the literature the disparity is usually estimated from the images I_1 and I_2 and the most commonly used data cost function is

$$\begin{aligned} E_d &= \|I_2(x, y) - I_1(x + D(x, y), y)\|_2 \\ &= \sum_{k=1}^{N_1} \|I_{2,k}^T - A^k I_{1,k}^T\|_2 \end{aligned} \quad (17)$$

where Eq. (17) follows from Eq. (4). In our work we are interested in estimating the disparity directly from the linear measurements. In the previous section we have shown that if the images are related linearly then the relationship between the corresponding measurements computed from the images remains linear as the sensing matrix is linear. Using Eq. (4) and Eq. (12) one can show that the data cost function relating the measurements Y_1 and Y_2 as,

$$\begin{aligned} E_d &= \sum_{k=1}^{N_1} \|I_{2,k}^T - A^k I_{1,k}^T\|_2 \\ &\approx \sum_{k=1}^{N_1} \|Y_{2,k} - \phi_2^k A^k \phi_1^{k\dagger} Y_{1,k}\|_2. \end{aligned} \quad (18)$$

Combining Eq. (18) and Eq. (15) the proposed energy model is given as,

$$E = \sum_{k=1}^{N_1} \|Y_{2,k} - \phi_2^k A^k \phi_1^{k\dagger} Y_{1,k}\|_2 + \lambda \sum_{\mathbf{z}, \mathbf{z}' \in \mathcal{N}} V_{\mathbf{z}, \mathbf{z}'}. \quad (19)$$

The dense disparity image D is estimated by minimizing the energy in Eq. (19), i.e., solving the optimization problem in Eq. (14). Due to the non-convexity nature of this problem the solution to Eq. (14) can be typically found by using strong optimization techniques based on graph-cuts or belief propagation.

4. EXPERIMENTAL RESULTS

We evaluate the performance of our scheme on two natural datasets, Tsukuba and Venus¹. The random projections are computed using a

¹available in <http://vision.middlebury.edu/stereo/data/scenes2001/>

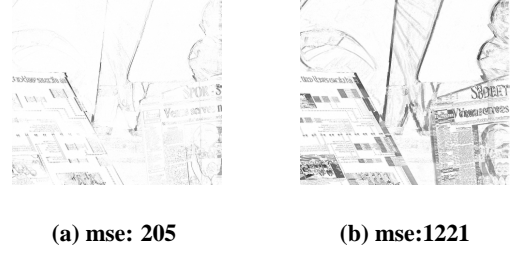


Fig. 2. PSNR comparison of the predicted image \hat{I}_2 w.r.t. I_2 and I_1 at a measurement rate 0.2. The disparity map shown in Fig. 1(a) is used to predict the image \hat{I}_2 . (a) $1 - |\hat{I}_2 - I_2|$ (b) $1 - |\hat{I}_2 - I_1|$. The error is inverted so that the white pixels correspond to no error.

scrambled Fourier transform where the scrambled operator is a diagonal matrix with entries ± 1 (taken from an i.i.d. Bernoulli random variable) [8]. We sampled both the images using the same measurement rate. The measurement rate is defined as the ratio between the number of computed measurements MN_1 and the dimensions of the image N_1N_2 . The disparity map D is estimated by solving Eq. (14) using α -expansion mode in graph-cuts [9, 10].

We first show the performance of the scheme when the measurement matrixes are different for both images. Fig. 1(a) shows the computed disparity map from a measurement rate 0.2 from each image (total rate is 0.4) for Venus dataset. Comparing with the ground truth image we observe that the result in Fig. 1(a) corresponds to a coarse approximation of the ground truth disparity image. Quantitatively the disparity error w.r.t. the ground truth is found out to be 41% when measured as the percentage of pixels with absolute error greater than one. When the measurement rate increases the quality of the disparity map is better as expected. For example, the estimated disparity map for a measurement rate 0.7 is shown in Fig. 1(b), and the corresponding disparity error is found to be 10%. When such a coarse disparity map D is used for image prediction the predicted image \hat{I}_2 is closer to I_2 than I_1 (see Fig. 2). We see in Fig. 2(a) that the prediction is accurate in the low frequency and smooth regions, and the prediction error is oriented mostly along the edges.

Fig. 3 shows the PSNR of the warped image \hat{I}_2 at various measurement rates for both datasets. From the plot we see that the quality of the predicted image \hat{I}_2 increases w.r.t. the measurement rate as it provides more information to estimate the disparity image. Furthermore the quality of the predicted image saturates at rates > 0.5 . It indicates that the quality of the disparity map reaches the saturation point and that cannot be further improved. In other words, the texture and high frequency components cannot be efficiently predicted by disparity compensation. While carrying out the experiments we also notice that the quality of the predicted image \hat{I}_2 is improved (by 6 - 8 dB on the datasets) by activating the smoothness cost in the optimization problem (i.e., $\lambda \neq 0$ in Eq. (19)). We then carry out experiments using a same set of measurement matrix. Fig. 3 compares the PSNR quality of the predicted image \hat{I}_2 w.r.t. the one predicted using different sets of measurement matrixes. It is clear that the prediction accuracy improves when different matrixes are used, as this brings more information from both images in order to solve the correspondence problem.

We then compare our results to a scheme that first reconstructs the images independently before estimating the disparity map. The images are reconstructed independently from the corresponding measurements by solving a convex optimization problem. We de-

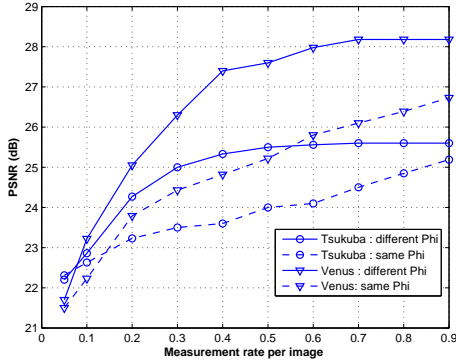


Fig. 3. Evolution of the quality of \hat{I}_2 at various measurement rate. The benefit of using different sets of measurement matrixes is also illustrated.

note this methodology as disparity from reconstructed images (DFR) in this paper. We tried out two different reconstruction methodologies. (1) DFR-sparsity: minimizing the l_1 norm of the sparse coefficients assuming that the image is sparse in a particular orthonormal basis (e.g., Wavelet). This problem is solved using GPSR [11]; (2) DFR-TV: minimizing the TV norm of the reconstructed images (solved using $l_1 - Magic^2$). After reconstructing the images the disparity map is estimated by solving Eq. (14) with the data cost and smoothness cost given in Eq. (17) and Eq. (15) respectively. Fig. 4 shows the comparison of the proposed scheme w.r.t. DFR-sparsity and DFR-TV schemes for the Tsukuba dataset. It is clear that we match the image prediction quality of the DFR-sparsity scheme (see Fig. 4). 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 outperforms (max. 3 dB) our scheme at low rates due to good image reconstruction quality. However, the image prediction quality is almost the same for all the three schemes at rates > 0.5 . We observe similar results on the Venus dataset. It should be noted that computational cost to estimate the correlation model directly from the measurements or from the references (reconstructed) images is approximately the same. Therefore when compared to the DFR scheme our scheme saves the image reconstruction cost; this corresponds to the cost needed to solve the $l_2 - l_1$ or $l_2 - TV$ optimization problem. Thus the proposed scheme significantly reduces overall complexity at the joint decoder when compared to the DFR schemes. In addition to lower complexity it is shown to work reasonably well in the low rate and as accurate as the DFR-TV scheme at high rate.

5. CONCLUSIONS

In this paper we propose a scheme to estimate the disparity image between a pair of images that are directly acquired based on random projections. We show that the disparity between the images can be efficiently represented using a linear operator and using this we derive a regularized cost function to estimate the disparity image directly in the compressed domain. Experimental results show that the proposed scheme gives a good estimation of the disparity image, especially when the images are sampled with different measurement matrixes. Finally we show that the performance of the proposed scheme does not degrade significantly w.r.t. the DFR-sparsity

²available in <http://www.acm.caltech.edu/l1magic/>

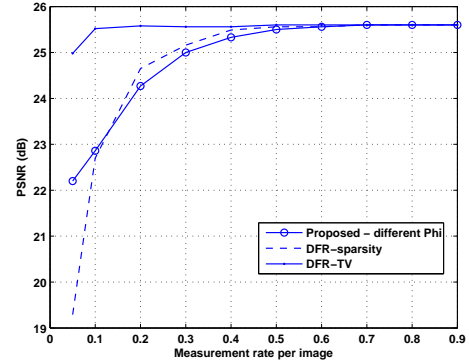


Fig. 4. Tsukuba dataset: Comparison of the predicted image quality \hat{I}_2 between the proposed, DFR-sparsity and DFR-TV schemes.

scheme. Thus the proposed scheme builds the correlation model directly in the compressed domain without any intermediate image reconstruction and thereby reduces the overall complexity at the decoder comparing to the DFR schemes.

6. 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] J. P. Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. PCS*, 2009.
- [6] J. M. Trocan, T. Maugey, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Compressed Sensing of Multiview Images Using Disparity Compensation", in *Proc. IEEE ICIP*, 2010.
- [7] V. Thirumalai and P. Frossard, "Distributed Representation of Geometrically Correlated Images with Compressed Linear Measurements," *submitted to IEEE Trans. on Image Proc.* [Online]. Available: <http://infoscience.epfl.ch/record/161685/files/>
- [8] L. Gan, "Block compressed sensing of natural images," in *Proc. Intl. Conf. DSP*, 2007.
- [9] O. Veksler, "Efficient graph based energy minimization methods in computer vision," Ph.D. dissertation, Cornell University, 1999.
- [10] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 23, pp. 1222-1239, Nov. 2001.
- [11] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Jour. on Sel. Areas in Comm.*, vol. 1, pp. 586-597, Dec. 2007.