Optimal Optical Flow based Disparity Map Estimation for Lossless Stereo Image Coding

Amit Kumar K.C.[a], Rony Darazi[a] and Benoît Macq[a]

[a]Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), Université catholique de Louvain
2 Place du Levant, B-1348, Louvain-la-Neuve, Belgium

ABSTRACT

In dependent stereo image compression, the aim is to minimize the bitrate of disparity map and that of residual image. Traditionally, focus has been paid on either disparity map or residual image. In this paper, we compute an optimal disparity map (in terms of bitrates) by jointly exploiting the trade-off between the disparity map and the residual image. Firstly, the dense disparity map is obtained using existing optical flow technique. Secondly, the dense disparity map is quantized using a RD framework. Consequently, the resulting bitrate of the disparity map decreases significantly at the cost of a slight increase of the bitrate of the residual image. As a result, the overall bitrate attains minimum value. The proposed scheme is compatible and can be integrated in JPEG2000 framework.

Keywords: Optical flow, Disparity map, Residual image, Block matching algorithm (BMA), JPEG2000 codec

1. INTRODUCTION

Stereo images are associted with the same scene, observed from two view points. The left and right views represent the stereo pair. In dependent stereo image coding, one image is considered as the reference image and other as the target image. The disparity between these two images is estimated and the residual image is obtained by computing the difference between the target image and the disparity compensated reference image. There exist a variety of methods to estimate the disparity map. Classically, block matching technique has been used to estimate the disparity. Another class of disparity map estimation is optical flow for which disparity is estimated at each point of the image assuming that the pixel intensity does not change on corresponding points.

According to Quentin *et al.* and Yu *et al.*,[1,2] the two classes, namely *optical flow* and *block matching*, are equivalent for subpixel displacements. Specifically, block matching that uses bilinear interpolation can be recast into equivalent optical flow formulations.

Since block matching and optical flow methods generate numerically equivalent estimates for subpixel displacements, other features may suggest choosing one over another. For example, the implementation of the optical flow algorithm may require significantly fewer operations than an equivalent block matching algorithm (BMA). The computational effort of a full search BMA is directly related to the size of the search window. In case of optical flow, there is no specific search window and the effective range is determined from characteristics of the image. In addition, the optical flow algorithm does not require any overhead to obtain subpixel disparity vectors as the interpolation is amortized within the gradient filters.[2]

Research in stereo image coding has focused on the disparity estimation and compensation process to exploit the inter-ocular (or cross-view) redundancies. The goal of stereo image coding is not to estimate the true disparity but rather to achieve a high compression ratio. Therefore, it may not be worthwhile to compute a dense disparity map if the cost of handling (transmitting or storing) the disparity vector field is too high. That is why, fixed size block matching has been widely used, even though the true disparity maps are blockwise constant.[3] Most of the reported methods use a classical block-based technique in order to estimate the disparity map. However, this

Further author information: (Send correspondence to Amit Kumar K.C.)
A.K.C.: E-mail: amit.kc@uclouvain.be

estimation technique does not always provide an accurate disparity map, which may affect the compensation step. Specifically, the block-based technique fails at object edges and boundaries. Besides, if the block size becomes smaller, the overhead required to transmit the disparity information becomes large with respect to overall bit rate.

Regarding disparity compensation, dense disparity map is superior but at the price of huge transmission/storage cost. However, the dense disparity map provides a great flexibility to tradeoff between the disparity map and the resulting residual image bitrates. The disparity map is computed only once. Then, one can set the bitrate for disparity map and compute the corresponding residual image. Since our aim is to minimize the sum of bitrates of disparity map and residual image, it allows an optimization on the accuracy of the disparity map with respect to required bitrate for the residual image. This is not possible for block-matching as the disparity map has to be computed every time for different block-size.

Morvan *et al.*[4] have used rate-distoriton(RD) framework based quadtree decomposition for depth map. In order to reduce bpp cost for disparity map, disparity map can be segmented using quadtree decomposion followed by median filtering as proposed by Kaaniche et al.[5] The idea behind[5] is to divide the estimated field into macro blocks. If the block is homogeneous, then it is kept intact, else it is recursively divided into four sub blocks. Thus, homogeneous areas will be represented by larger blocks whereas boundaries will have smaller blocks.

However, above-mentioned works consider either the residual image or the disparity map but not both for quantization. Thus, these techniques are difficult to adopt for lossless encoding which requires both the disparity map and the residual image to be considered for encoding. In this paper, the goal is to compute the optimal disparity map (in terms of bitrates) by jointly exploiting the trade-off between the disparity map and the residual image. The novelty of this work is that both the disparity map and the residual image are jointly considered and consequently the optimum quantization of the disparity map is done such that the overall bitrate is minimized.

The organization of this paper is as follows. In section 2, the proposed optimization scheme is presented. Experimental results are shown in section 3, followed by conclusion in section 4.

2. DISPARITY MAP ESTIMATION AND OPTIMAL QUANTIZATION

2.1 Disparity map estimation

In the optical flow method, the disparity map estimation problem can be stated as a minimization problem which minimizes the following energy function.[6]

$$J = \int |\nabla \underline{u}| d\underline{x} + \lambda \int |I_L(\underline{x} + \underline{u}(\underline{x}) - I_R(\underline{x})| d\underline{x} \tag{1}$$

where ∇u is the gradient of the disparity map u, the data term is represented by the L_1 norm of the difference between the left image I_L and the right image I_R, the regularization term is represented by total variation (TV) norm of u. Assuming only the horizontal disparity such that $\underline{u}(\underline{x}) = u(\underline{x})$ and linearizing $I_L(\underline{x} + u(\underline{x}))$ around initial point $x + u_0(\underline{x})$, we can write the data term of euqation 1 as:

$$I_L(x + u(\underline{x})) = I_L(x + u_0(\underline{x})) + (u(\underline{x}) - u_0(\underline{x})) I_L^x(x + u_0(\underline{x})) \tag{2}$$

where u_0 is the given disparity map and I_L^x is the horizontal derivative of I_L. For simplicity, we write $u \equiv u(\underline{x})$ and $x \equiv \underline{x}$.

The above problem is reduced to following equation:

$$\begin{aligned} J &= \int |\nabla u| dx + \lambda \int |u I_L^x(x + u_0) + I_L(x + u_0) - u_0 I_L^x(x + u_0) - I_R(x)| \, dx \\ &= \int |\nabla u| dx + \lambda \int |\rho(u)| dx \end{aligned} \tag{3}$$

where $\rho(u) = u.I_L^x(x + u_0) - u_0.I_L^x(x + u_0) + I_L(x + u_0) - I_R(x)$.

To solve the above optimization problem, we introduce an auxiliary variable v and rewrite the above equation 3 as:

$$J_\theta = \int \left[\lambda|\rho(u)| + \frac{1}{2\theta}(u-v)^2 + |\nabla u| \right] dx \tag{4}$$

where θ is a small constant such that v is close approximation of u.

Thus, we can decouple the above minimization problem of equation 4 into two problems:

- For fixed v, solve
$$\min_u \int \left[|\nabla u| + \frac{1}{2\theta}(u-v)^2 \right] dx \tag{5}$$
This is solved by using *Chambolle's approach*[7] and the solution is given by $u = v - \theta div(\underline{p})$, where $div(.)$ is the divergence and $\underline{p}$ is obtained iteratively as
$$\underline{p}^{k+1} = \frac{\underline{p}^k + \tau \nabla \left(div(\underline{p}^k) - \frac{v}{\theta} \right)}{1 + \tau \left| \nabla \left(div(\underline{p}^k) - \frac{v}{\theta} \right) \right|} \tag{6}$$
with $\underline{p}^0 = \underline{0}$ and time step $\tau \leq 1/8$

- For fixed u, solve
$$\min_v \int \left[\lambda|\rho(v)| + \frac{1}{2\theta}(u-v)^2 \right] dx \tag{7}$$
The solution is
$$v = u + \begin{cases} \lambda\theta I_L^x & \text{if } \rho(u) \leq -\lambda\theta \left(I_L^x\right)^2 \\ -\lambda\theta I_L^x & \text{if } \rho(u) > -\lambda\theta \left(I_L^x\right)^2 \\ -\frac{\rho(u)}{I_L^x} & \text{otherwise} \end{cases} \tag{8}$$

Equation 8 can be equivalently considered as soft-thresholding of u to obtain v.

Since the original problem is non-convex function and is transformed into convex problem by linearization, the algorithm is valid only for small displacements and hence the energy minimization procedure is embedded into a coarse-to-fine approach. The pyramid is constructed with a down-sampling factor of 2. Beginning with coarsest level, we solve the algorithm and the results are propagated downwards the pyramid.

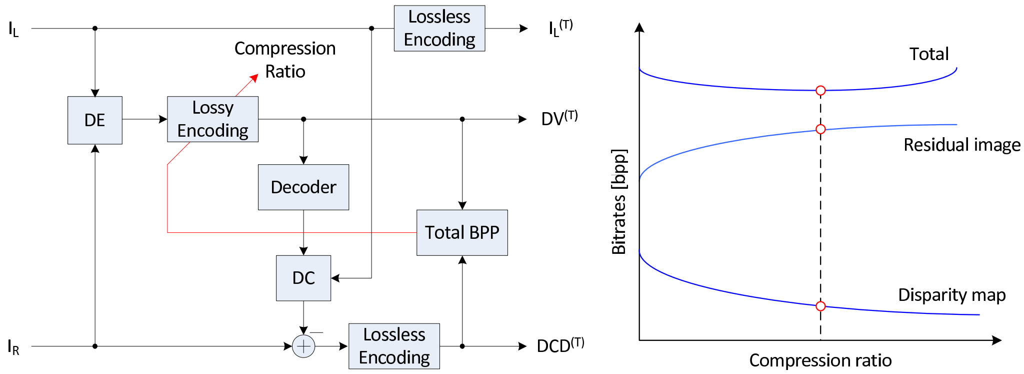
Since the dimensions of sub-bands at resolution level j are half of those at level $j-1$, the disparity vectors are multiplied by a factor of 2 while propagating downwards the pyramid. The coarse-to-fine approach not only accelerates the convergence, but also avoids local minima. It is important to note that the above equation may yield non integer values of the disparity vectors. In that case, the corresponding right image pixel is obtained by using bilinear interpolation.

$$u_{j-1}(x,y) = 2u_j(2x,2y) \qquad \text{for } j = N \text{ to } 1 \tag{9}$$

The Lagrangian parameter λ is used to determine which term (data or regularization) should be given more weight during optimization process. In other words, it balances between the data fidelity and the regularization terms. Larger value of λ exhibits more priority to the data term and hence the disparity compensated difference (DCD) is improved. Thus, we expect that the entropy of the DCD decreases with increasing λ. On the other hand, smaller value of λ exhibits more priority to the disparity map and hence the entropy of the disparity map is reduced. Therefore, the free parameter λ shows a tradeoff between the accuracy of the DCD and the disparity map.

2.2 Optimal quantization of disparity map

Once the dense disparity map is obtained, an optimal quantization is performed on the dense disparity map such that the total bitrate of resulting disparity map and the residual image is minimized. The block diagram is presented in figure 1:



(a) Proposed scheme (b) Conceptual diagram

Figure 1. Proposed scheme for optimal quantization of disparity map

We quantize disparity map with different compression ratios using JPEG2000 codec as discussed earlier. It is obvious that as we quantize the disparity map, the residual image quality gets worsened i.e., the bitrate for residual image $R^{(e)}$ increases, whereas the bitrate for the disparity map $R^{(v)}$ decreases. It is worth noting the rate at which $R^{(v)}$ changes is different than the rate at which $R^{(e)}$ changes. Specifically, $R^{(v)}$ decreases much rapidly and $R^{(e)}$ increases slowly until some quantization level. Beyond this level, $R^{(v)}$ decreases slowly whereas $R^{(e)}$ starts increasing sharply. Since we are interested at the total bpp (i.e., $R^{(v)}+R^{(e)}$), the total bitrate exhibits a minimum. This can be illustrated in figure 1. It can be easily observed that our proposed scheme allows to have a direct tradeoff between the bitrate of disparity map and that of residual image. The advantages of the proposed method as compared to quadtree decomposition method(QTD)[5] are:

- QTD technique requires side information about the block size and segmented disparity map to be sent. This is not required in the proposed scheme.

- QTD is not directly compatible to JPEG2000 as it incorporates differential pulse code modulation (DPCM) encoding for transmission of encoded bits. Our method uses the JPEG2000 encoder to encode the disparity map as normal image, thereby alleviating the need of using DPCM. The proposed method is an extension of our previous work which deals with compression of stereo images for digital cinema using directional and optimized prediction techniques.[8]

3. EXPERIMENTAL RESULTS

For the simulation, stereo images are downloaded from Carnegie Mellon and Middlebury websites.[9–11] Figure 2 depicts the qualitative behaviour of the disparity map and the residual image for different compression ratios in the case of *Tsukuba* image.

From the above figure, it can be noted that the quality of the disaprity map does not change much even for high compression ratio. The reason behind it is that the disparity map (obtained by using total variation as regularization) has cartoon shape, i.e., it has almost smooth regions separated by edges. One possible drawback of compressing disparity map at very high compression ratio is appearance of spurious edges in the disparity map or disappearance of some weak edges.

Figure 3 verifies our assumption about the resulting bitrates (bpp curves) for the residual image and the disparity map. It depicts the behaviour of the disparity map and the residual image for various compression ratios (and hence various bitrates). Moreover, the reconstruction quality of the disparity map is also presented

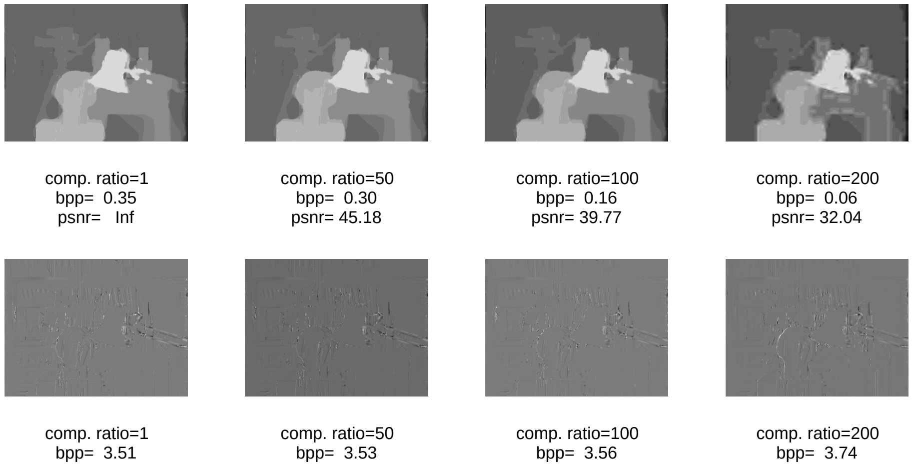
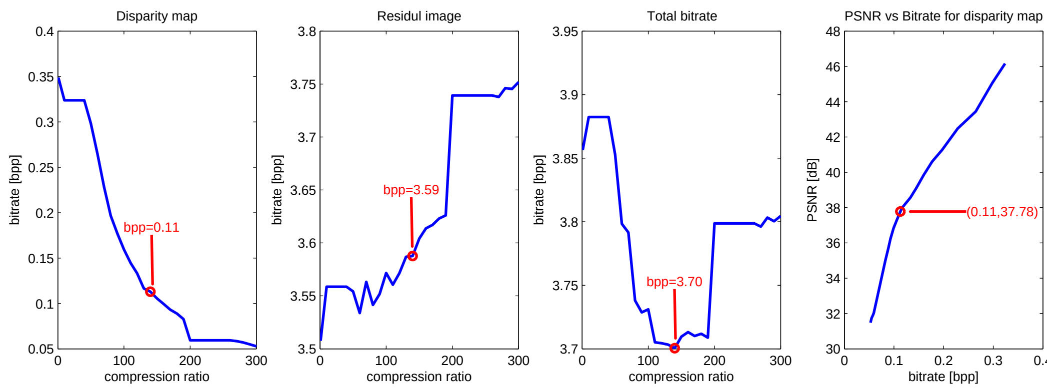
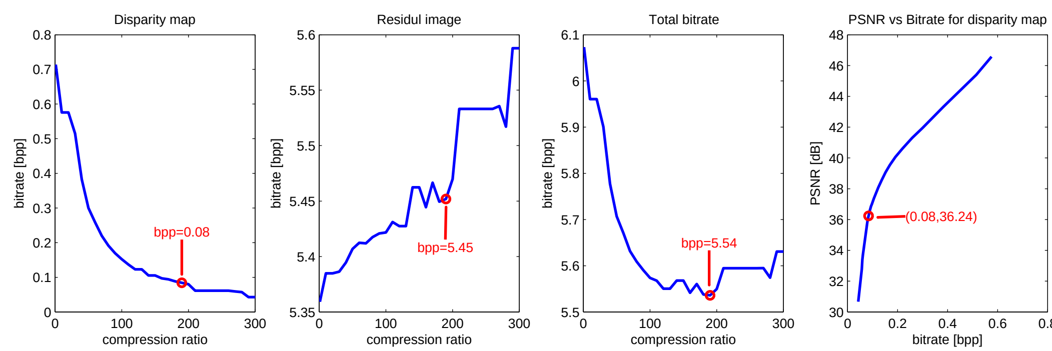


Figure 2. Quality of disparity map and residual image



(a) Tsukuba



(b) HouseOf

Figure 3. Bitrates and PSNR for *Tsukuba* and *HouseOf* images

in terms of peak-signal-to-noise-ratio (PSNR) between the unquantized disparity map,u, and the reconstructed disparity map after quantization, $\tilde{u}$. Table 1 gives numerical results of above-mentioned optimization procedure. It depicts the bitrates of the disparity map, the residual image and the total rate before and after optimization for different images. It can be easily concluded that the optimization procedure gives an advantage of about 0.3 bpp on total bitrate. We can observe from the table 1 that the bitrate for the disparity map drops drastically from an average of 0.56 bpp to 0.07 bpp. On the other hand, the increase of the bitrate of the residual image bpp is not very significant. It increases from 4.04 bpp to 4.13 bpp on average.

Table 1. Results of optimization of dense disparity map

Image	**Bitrates before optimization**			**Bitrates after optimization**		
	Disparity	**Residue**	**Total**	**Disparity**	**Residue**	**Total**
Tsukuba	0.35	3.51	3.86	0.11	3.59	3.70
Fruit	0.49	4.05	4.54	0.06	4.10	4.16
Pentagon	0.64	5.11	5.75	0.06	5.21	5.27
Apple	0.87	4.54	5.41	0.10	4.72	4.82
HouseOf	0.71	5.36	6.07	0.08	5.45	5.53
Corridor	0.28	2.05	2.33	0.06	2.09	2.15
Pm	0.10	3.27	3.37	0.05	3.27	3.32
Birch	0.97	4.56	5.53	0.05	4.72	4.77
Shrub	0.10	3.27	3.37	0.05	3.27	3.32
Average	0.56	4.04	4.59	0.07	4.13	4.20

4. CONCLUSION

In this paper, an optical flow based disparity estimation method for stereo images has been presented. For estimation of the disparity map, $TV - L_1$ based approach has been used as *total variation* based method preserves the edges in disparity map and is quite robust to outliers and noise. Although the dense disparity map is not suitable for direct coding, it offers flexibility to be encoded such that the total bitrate of the disparity map and the residual image is minimized. In addition, the proposed method offers a rate-distortion framework for optimal quantization of the disparity map without any prior segmentation. So, the disparity image is compressed as a normal image and is compatible with JPEG2000 environment. Results in terms of bpp show an improvement of about 0.4 bpp on average.

5. ACKNOWLEDGEMENTS

This work is partly supported by BCRYPT project, a Belgian Interuniversity Attraction Pole IAP-VI fund programme and the Walloone Region of Belgium.

REFERENCES

[1] C. Davis, Z. Karul, and D. Freeman, "Equivalence of subpixel motion estimators based on optical flow and block matching," in *Computer Vision, 1995. Proceedings., International Symposium on*, pp. 7 –12, 21-23 1995.

[2] Y. Chi, T. Tran, and R. Etienne-Cummings, "Optical flow approximation of sub-pixel accurate block matching for video coding," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, **1**, pp. I–1017 –I–1020, 15-20 2007.

[3] W. Woo and A. Ortega, "Stereo image compression with disparity compensation using the mrf model," in *Proc. SPIE VCIP*, pp. 28–41, 1996.

[4] Y. Morvan, D. Farin, and P. de With, "Depth-image compression based on an r-d optimized quadtree decomposition for the transmission of multiview images," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, **5**, pp. V –105 –V –108, sept. 2007.

[5] M. Kaaniche, W. Miled, B. Pesquet-Popescu, A. Benazza-Benyahia, and J.-C. Pesquet, "Dense disparity map representations for stereo image coding," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 725 –728, 7-10 2009.

[6] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l1 optical flow," in *Pattern Recognition (Proc. DAGM)*, pp. 214–223, (Heidelberg, Germany), 2007.

[7] A. Chambolle, "An algorithm for total vairation minimization and applications," *Journal of Mathematical Imaging and Vision* , pp. 89–97, 2004.

[8] R. Darazi, A. Gouze, and B. Macq, “Adaptive lifting scheme-based method for joint coding 3d-stereo images with luminance correction and optimized prediction,” *Acoustics, Speech, and Signal Processing, IEEE International Conference on* **0**, pp. 917–920, 2009.

[9] “http://vasc.ri.cmu.edu/idb/html/stereo/index.html.”

[10] “http://vasc.ri.cmu.edu/idb/html/jisct/index.html.”

[11] “http://vision.middlebury.edu/stereo/data/scenes2001/.”