

Supplementary Material: A Convex Solution to Disparity Estimation from Light Fields via the Primal-Dual Method

Mahdad Hosseini Kamal¹, Paolo Favaro², and Pierre Vandergheynst¹

¹ Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland,
{mahdad.hosseini.kamal,pierre.vandergheynst}@epfl.ch

² University of Bern, Bern 3012, Switzerland
paolo.favaro@iam.unibe.ch

1 Preconditioning

The convergence speed of the primal-dual algorithm depends on the steps σ and τ . These steps must satisfy the constraint $\tau\sigma\|K\|_2^2 < 1$, i.e the operator norm of K influences on the convergence speed of primal-dual. Given the current form of K , the operator norm of k depends on the operator norm of the patch dictionary Q . The patch dictionary is a dense block diagonal matrix therefore its operator norm, therefore the possible values of σ and τ are unfavorable. To improve the convergence rate we employ customized per-row and per-column steps τ_d and σ_c and, as suggested in [1], set them according to

$$\begin{aligned}\tau_d &= \frac{1}{\sum_{c=1}^{\tilde{M}\tilde{N}W^2(2+3D)-(\tilde{M}+\tilde{N})DW^2} |K_{cd}|} \\ \sigma_c &= \frac{1}{\sum_{d=1}^{\tilde{M}\tilde{N}((nm-1)D+W^2)} |K_{cd}|}\end{aligned}\tag{1}$$

where σ_c is divided into different parts for F_i s. For F_1 , is defined as $\omega_{x,y} = \sum_k \frac{1}{\|Q_{x,y}^k\|}$ and for F_3 is equal to $\xi_{x,y} = 2$ if (x,y) belongs to the domain Ω except for the bottom row and the right column, and $\xi_{x,y} = 1$ if (x,y) belongs to the bottom row and the right column except the bottom-right corner. We employ Moreau's identity as explained in the paper to calculate the proximity operators for the functions.

2 Implementation Details

2.1 Convex labeling

Given C , the disparity $\hat{\rho}$ is determined at each pixel (x,y) independently by solving

$$\hat{\rho}_{x,y} = \arg \max_{\rho \in \{\rho_1, \dots, \rho_D\}} \|C_{x,y}^\rho\|.\tag{2}$$

However, one can introduce a convex labeling by fitting a convex function per patch to the estimated $C_{x,y}$ and impose a smooth prior such as TV constraint on the recovered disparity to solve an inpainting problem.

$$\tilde{\rho} = \arg \min_{\rho} \lambda \|\nabla \rho\|_2 - f_C(\rho), \quad (3)$$

where $\lambda > 0$ is a constant and f_C is the convex function fitted to $C_{x,y}$. This formulation can help to remove the disparity noise however it can smooth some details.

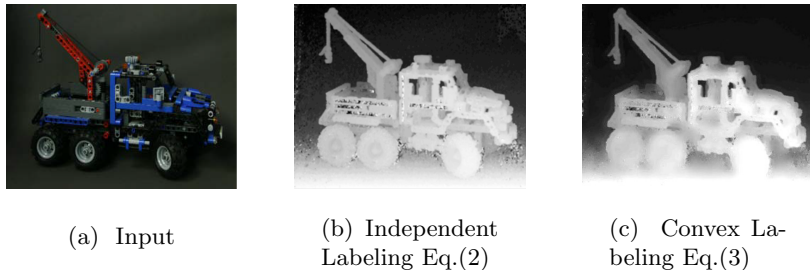


Fig. 1. Comparison of smooth disparity estimation with independent disparity estimation per pixel. We observe that smooth labeling can remove some noise as the result of improper initial candidate selection (specially in smooth area), however it smooths out some details with respect to independent disparity labeling.

2.2 Improving initial disparity candidate selection

One shortcoming of our method is that it is very computationally intensive. Because of that we had to limit the number of possible depth levels. To do so we use the plane sweep method and at each pixel we compute a matching error for each depth hypothesis (see sec. 6.4). While this strategy allows us to reduce the number of possible depth levels and thus make our algorithm run much faster, when none of the depth candidates is close to the correct solution the final estimate will be noisy. We can see this effect in several recovered depth maps. One way to improve the final estimate is clearly to improve the candidates by using larger patches, or more views, or by using smarter heuristics to choose candidates. Another strategy that we will explore in the future is simply to avoid the depth reconstruction at pixels where we do not have good candidates. Hence, we can estimate C on a subset Ψ of the set of all image pixels. One can then recover depth at every pixel later on given the estimated C . If C is not defined at all pixels, one can then solve the inpainting problem

$$\tilde{\rho} = \arg \max_{\rho} \sum_{x,y \in \Psi} (\rho_{x,y} - \hat{\rho}_{x,y})^2 + \varepsilon \|\nabla \rho\|_2 \quad (4)$$

where $\varepsilon > 0$ is a constant. Finally, we will explore the Block Coordinate Descent (BCD) [2] in detail on our algorithm, where we simply iterate on subsets of coefficients at each time. This way we can work on the full depth candidates at each time to avoid uncertain disparity map. Fig. 2 demonstrates disparity estimation using BCD on 100 disparity candidates without pre-selection of the most proper candidates. The disparity labeling from the estimated C is performed both by the convex and the independent method. We observe that the convex model leads to a smoother result however the independent labeling using Eq. (2) has more details.

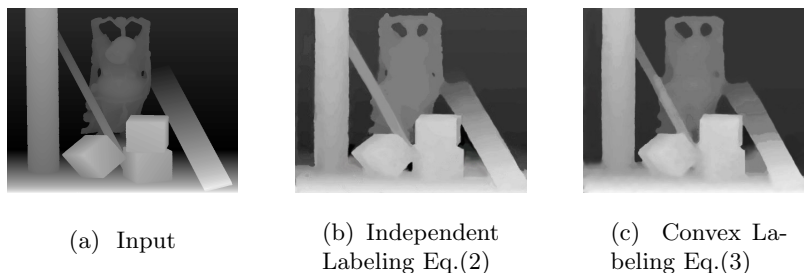


Fig. 2. Estimating disparity map with 100 disparity candidate using block coordinate descent method.

3 Multiview Depth Estimation

We test our light field depth estimation technique on the Middlebury dataset³ with stereo methods [3], two light field depth estimation schemes [4, 5], and convex formulations [1, 6]. Our parameters are: $\mu = 0.6$ and $\gamma = 1$ for all datasets, then $\lambda = 0.5$ for Venus, $\lambda = 0.15$ for Cone.

One aspect we would like to point out is that the number of views used in the depth estimation problem improves the depth estimate considerably, this is clearly noticeable in Fig. 3. Our algorithm is also demonstrated in the limit case where there are only two views (Fig. 4). The group sparsity constraint can still work quite successfully.

References

1. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences* **3**(4) (2010) 1122–1145

³ <http://vision.middlebury.edu/stereo/data/>

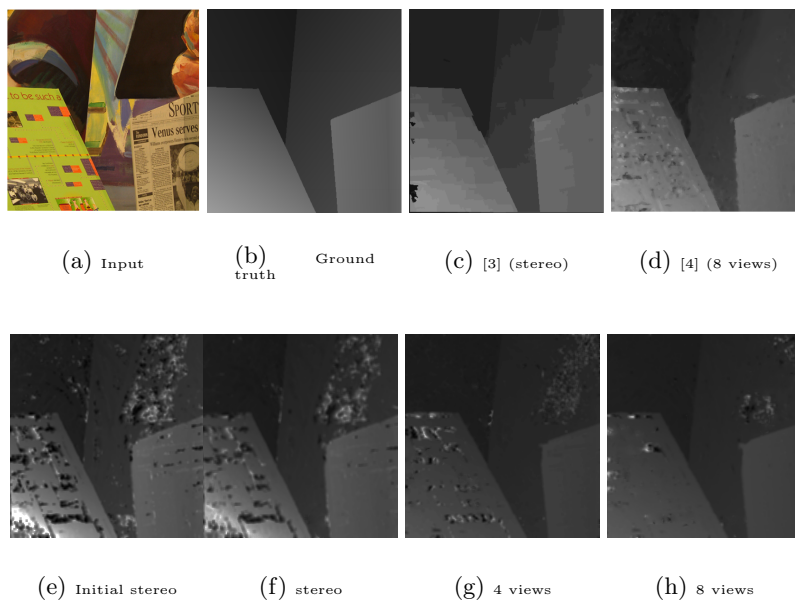


Fig. 3. Venus dataset. On the top row we show (left to right): one input image, the ground truth depth map, the estimate of [3] for the stereo case and that of [4] for 8 views. On the bottom row we show (left to right): our initial depth estimate (plane sweep depth search), our final result with 2, 4 and 8 views.

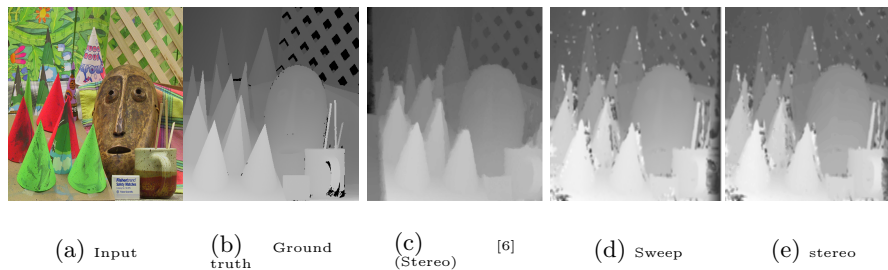


Fig. 4. Cone dataset. Top image is one of the input images. Bottom row (left to right): the ground truth depth map, the estimate of [6] for the stereo case, our initial depth estimate (plane sweep depth search), and our final result with 2 views.

2. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications* **109**(3) (2001) 475–494
3. Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video view interpolation using a layered representation. In: *ACM Transactions on*

- Graphics (TOG). Volume 23., ACM (2004) 600–608
4. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.: Scene reconstruction from high spatio-angular resolution light fields. To appear ACM Trans. Graph.(Proc. SIGGRAPH) (2013)
 5. Wanner, S., Goldluecke, B.: Globally consistent depth labeling of 4d light fields. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 41–48
 6. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: Computer Vision–ECCV 2008. Springer (2008) 792–805