

Performance Evaluation of Narrow Band Methods for Variational Stereo Reconstruction

Franz Stangl, Mohamed Souiai and Daniel Cremers

TU Muenchen

Abstract. Convex relaxation techniques allow computing optimal or near-optimal solutions for a variety of multilabel problems in computer vision. Unfortunately, they are quite demanding in terms of memory and computation time making them unpractical for large-scale problems. In this paper, we systematically evaluate to what extent narrow band methods can be employed in order to improve the performance of variational multilabel optimization methods. We review variational methods, we present a narrow band formulation and demonstrate with a number of quantitative experiments that the narrow band formulation leads to a reduction in memory and computation time by orders of magnitude while preserving almost the same quality of results. In particular, we show that this formulation allows computing stereo depth maps for 6 Mpixels aerial image pairs on a single GPU in around one minute.

1 Introduction

1.1 Convex Multilabel Optimization and Narrow Band Methods

Most of the relevant algorithmic challenges in computer vision correspond to energy minimization problems with non-convex energies. While traditional approaches to segmentation [10], stereo [3] and optical flow [8] aimed at finding acceptable solutions by local minimization starting from an “appropriate” initialization, in the last few years researchers have proposed convex relaxation techniques which allow to compute optimal or near-optimal solutions [15, 14, 17, 11, 4, 6]. The key idea in these algorithms which was inspired by Ishikawa’s graph theoretic approach [9] is to increase the dimension of the optimization problem by enhancing the spatial dimensions with the label dimension. Specifically, it was shown in [15] that the non-convex stereo reconstruction problem in two spatial dimensions is equivalent to a convex optimization problem in three dimensions.

While bringing about a clear gain in optimality, this increased dimension comes with an important sacrifice in memory and computation since the size of these lifting methods increases linearly with the number of labels used for the reconstruction. As a consequence, these algorithms have limited practical use – even for smaller problems of around 640×480 pixel stereo reconstructions they can easily take a minute of computation time. Moreover, the direct application to large-scale aerial images of several megapixels is entirely infeasible due to memory limitations. A popular strategy to accelerate such algorithms and reduce their memory requirements is to apply coarse-to-fine narrow band

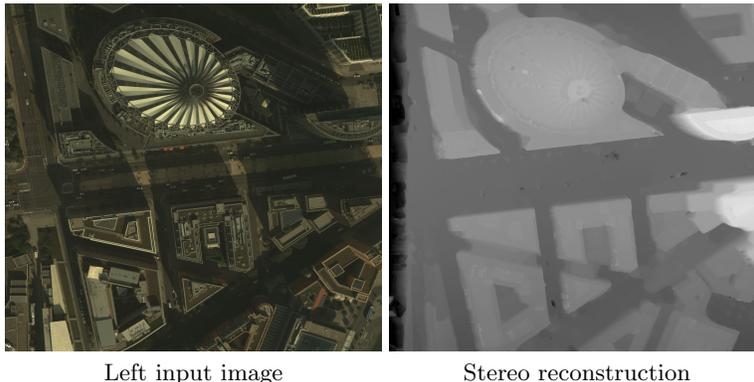


Fig. 1. A large scale reconstruction (right) on two **2600 x 2400** aerial images¹ (left) with **300** disparity levels is performed on a NVIDIA GTX 680 GPU with 4 GB of memory. Memory demand for reconstruction is **2.6 GB** with a runtime of **77 seconds**.

methods [1, 12, 2]. Since in these algorithms the optimal stereo reconstruction is given by the isolevel of an embedding function defined on the 3D grid, one can compute the embedding function in a coarse to fine manner, considering on each scale of the hierarchy a narrow band of a few levels around the current solution. While this strategy does not preserve optimality for the finest scale – small unconnected fine-scale structures potentially being lost – it allows to drastically reduce memory and computation time.

1.2 Contribution

While there are local methods (e.g. [7]) which can handle large scale stereo reconstruction, we intend in this paper to systematically evaluate the tradeoff between the loss in accuracy and the increase in speed and memory brought about by narrow band methods for variational multilabel optimization. More specifically, we focus on the problem of stereo reconstruction using a variational approach with non-convex data term and convex total variation regularizer. In Section 2, we briefly review the variational approach of Pock et al. [15] for globally optimal stereo reconstruction. In Section 3, we will present a narrow band formulation of this approach. In Section 4, we will present a detailed quantitative analysis of the performance gain brought about by the narrow band formulation. We also study how the accuracy increases with the width of the narrow band. In particular, we will show that near-optimal solutions can be computed with drastically reduced memory and computation time. We can recover high-resolution stereo depth reconstruction from 6 megapixels images in around one minute on a single end-user GPU. These experiments demonstrate that in combination with narrow band methods, convex relaxation techniques for multilabel optimization exhibit an enormous practical potential for highly accurate large-scale reconstruction. An example of a large scale reconstruction is given in Figure 1. While

¹ Image courtesy of Heiko Hirschmüller, German Aerospace Center (DLR) Institute of Robotics and Mechatronics.

the input image pair exceeds 6 megapixels in resolution, we were able to perform the reconstruction using 300 depth labels on a consumer graphics card. The memory requirement is around 2.6 GB and is a drastic gain compared to a reconstruction with dense label space that would require 43 GB of memory which no commercially available graphics cards can offer.

2 A Convex Formulation of Multi Label Stereo

2.1 Continuous Setting

In this work we devote ourselves to the study of how to efficiently minimize the following variational problem:

$$\min_u \left\{ \int_{\Omega} |Du(x)| + \int_{\Omega} \varrho(u(x), x) dx \right\}, \quad (1)$$

where $\Omega \subset \mathbb{R}^2$ denotes a continuous image domain, and $u : \Omega \rightarrow \Gamma$ an unknown function which maps each point in Ω to a real valued range $\Gamma := [t_0, t_{\text{end}}]$. The second term in equation (1) assigns a point-wise cost for each pixel taking on a certain value from Γ . The data term in our application can be arbitrary and not necessarily convex as we will see in the convex relaxation Section 2.2. In order to impose a spatial regularity while preserving its discontinuity we make use of the total variation of function u given by the left term of equation (1). Note that Du is the gradient in the distributional sense since function u must not be differentiable as in the case of natural images. The energy given in (1) can be considered as the continuous counterpart of the discrete setting described by Ishikawa et al [9].

2.2 Convex Relaxation

Although the total variation regularizer in energy (1) is a convex functional, many interesting problems from vision are associated with a non-convex data term. This makes computing the global minimizer of the energy almost impossible and the straightforward minimization of the functional above is prone to getting stuck in local minimizers. We next describe an approach given in [15] which tackles the problems shown above.

The functional lifting approach: The authors in [15] devise a level set formulation of energy (1):

$$\min_{v \in \mathcal{C}} \left\{ \int_{\Omega \times \Gamma} |\nabla_x v(x, t)| + \varrho(x, t) |\nabla_t v(x, t)| dx dt \right\}, \quad (2)$$

where ∇_x and ∇_t denote the spatial gradient in Ω and respectively the gradient with respect to the label dimension Γ and where

$$\mathcal{C} = \{v(x, t) : \Omega \times \Gamma \rightarrow \{0, 1\}, v(\cdot, t_0) = 1, v(\cdot, t_{\text{end}}) = 0\}. \quad (3)$$

This new formulation is now a functional of the function v which is the indicator function of the subgraph of u hence

$$v(x, t) = \begin{cases} 1 & \text{if } t \leq u(x) \\ 0 & \text{else.} \end{cases}$$

Energy (2) becomes convex if we relax the range of indicator function v to the unit interval $[0, 1]$. Hence we constrain v by the relaxed constraint

$$\tilde{\mathcal{C}} = \{v(x, t) : \Omega \times \Gamma \rightarrow [0, 1], v(\cdot, t_0) = 1, v(\cdot, t_{\text{end}}) = 0\}. \quad (4)$$

The non-differentiability of the total variation is tackled by using its dual formulation which gives us overall the following saddle-point problem:

$$\min_{v \in \tilde{\mathcal{C}}} \left\{ \sup_{\Phi \in \mathcal{K}} \int_{\Omega \times \Gamma} \Phi \cdot \nabla v \, dx dt \right\}, \quad (5)$$

with the dual variable $\Phi(x, t) = (\Phi^x(x, t), \Phi^t(x, t))^T$ constrained by the following convex set:

$$\mathcal{K} = \{\Phi(x, t) : \Omega \times \Delta \rightarrow \mathbb{R}^3, |\Phi^x(x, t)| \leq 1, |\Phi^t(x, t)| \leq \varrho(x, t)\}.$$

Note that the saddle-point formulation (5) is linear in both its dual variable Φ and its primal variable v and is endowed with a convex constraint set. This renders our problem solvable using so called primal dual algorithms.

2.3 Convex Optimization

In order to solve problem (5) we make use of a first order primal-dual algorithm devised in [5] which essentially performs a gradient ascent in the dual variable and a gradient descent in the primal variable with subsequent orthogonal projection onto the respective convex sets. For more details see [5].

In the following, we detail solving the saddle-point problem using the primal dual scheme. We initialise $((\tilde{v})^0, (\varphi)^0) \in \tilde{\mathcal{C}} \times \mathcal{K}$, let $(\hat{v})^0 = (\tilde{v})^0$ and choose the time-steps σ and τ according to a preconditioning scheme presented in [13]. Then the iterates of the primal dual algorithm can be written as follows:

$$\begin{cases} (\Phi)^{n+1} &= \text{proj}_{\mathcal{K}}((\Phi)^n + \sigma(\nabla \hat{v}^n)), \\ (\tilde{v})^{n+1} &= \text{proj}_{\tilde{\mathcal{C}}}((\tilde{v})^n + \tau(\text{div } \Phi^{n+1})), \\ (\hat{v})^{n+1} &= 2(\tilde{v})^{n+1} - (\tilde{v})^n, \end{cases} \quad (6)$$

where the discretized divergence operator div is chosen to be adjoint to the discretisation of the gradient ∇ . For the data term we choose a simple sum of the channel-wise absolute differences between the rectified right and left image for every disparity level $t \in \Gamma$ i.e.

$$\varrho(x, t) = \frac{1}{3} \lambda \sum_{i \in \{r, g, b\}} |I_L^i(x) - I_R^i(x + (0, t)^T)|.$$

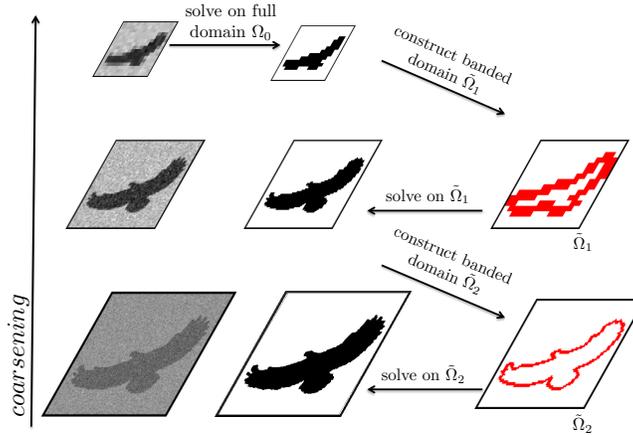


Fig. 2. An illustration of the narrow band method for 2 scales on a binary segmentation example. After coarsening the input image twice, a solution on the coarsest scale is computed. Using this solution the narrow band (red contour) of the next fine scale is computed and we perform the optimization solely on that region.

3 Narrow Band Formulation of Multilabel Optimization

While we are able to compute the global minimizer of energy (1), we pay the price of adding an additional dimension to the optimization problem. The increased amount of variables increases the runtime and the memory usage of our optimization algorithm tremendously. This makes the functional lifting approach scale bad with increasing resolution of the input images. Additionally, it becomes difficult computing depth maps of even moderate image sizes on consumer GPU's because of their limited memory. For example the dense reconstruction of a 6 megapixels image using 300 disparity levels, would require 43 GB of GPU memory which cannot be found on any graphics card on the market. In the following, we elaborate on the so called narrow band method which is a promising approach for leveraging these problems – see Figure 2.

3.1 The Narrow Band Idea

The basic idea of the narrow band method is to compute a solution on a coarse scale and use a narrow band around the 0-1 interface of its up-sampled version to construct a solution on the fine scale. This procedure can also be done for multiple scales where we propagate the computed solution via up sampling until we reach the original resolution. The coarsening process is done by resizing the two rectified images at each scale to half of their width and height. We use these down-sampled stereo pairs to compute the data term on each scale for the voxels that are represented by the NB and solve equation (1) by means of algorithm (6). Overall we perform the following steps for a NB method with K scales :

1. Construct an image pyramid of the input rectified image pairs and calculate a dense data term only for the lowest scale.

2. Compute solution v_0^* on the coarsest scale.
3. For $k : 1 \cdots K - 1$ do the following: Construct a banded domain $\hat{\Omega}_k$ using the neighbourhood of the 0-1 interface of the rounded solution v_{k-1}^* (for details see next section) and compute v_k^* on $\hat{\Omega}_k$.

An illustration of above algorithm can be found in Figure 2. Next section details the computation of the banded domain $\hat{\Omega}$.

3.2 Creating the Narrow Band

In the following we denote the discretisation of the domain by the superscript h . Creating an efficient representation for the narrow band in the domain $\Omega^h \times \Gamma^h$ is not straightforward. This is due to the loss of the natural grid indexing in the domain $\Omega^h \times \Gamma^h$ which requires a new strategy. We create the banded domain $\hat{\Omega}_{k+1}$ by the following 4 steps:

1. We transform the voxel ($k = 0$) and respectively the narrow band ($k \geq 1$) representation of the rounded solution v_k^* into a two dimensional array $I_k : \Omega_k^h \rightarrow \Gamma_k^h$ which holds for each pixel the individual label number.
2. Upsampling by factor of 2 from I_k to I_{k+1} by doubling the width and the height of the image domain as well as the label range.
3. Narrow Band creation: We implicitly represent the narrow band by creating an upper $U_{k+1} : \Omega_{k+1}^h \rightarrow \Gamma_{k+1}^h$ and lower bound $L_{k+1} : \Omega_{k+1}^h \rightarrow \Gamma_{k+1}^h$ on the voxel lying in the narrow band as follows:

$$L_{k+1}(x, y) = \min_{(v,w) \in B(x,y)} (I_k(v, w)) - \frac{NBT}{2} + 1$$

$$U_{k+1}(x, y) = \max_{(v,w) \in B(x,y)} (I_k(v, w)) + \frac{NBT}{2}$$

where $B(x, y)$ denotes a ball around pixel (x, y) and NBT is the chosen *Narrow Band Thickness*. We chose B to be an L^1 ball with the NBT as the diameter hence:

$$B(x, y) = \left\{ v, w \mid |v - x| + |w - y| \leq \frac{NBT}{2} \right\}.$$

We use $\min_{(v,w) \in B(x,y)} (I_k(v, w))$ and $\max_{(v,w) \in B(x,y)} (I_k(v, w))$ in order to account for the biggest jumps in the neighbourhood $B(x, y)$ and include these in the narrow band.

4. Now that we have located the banded domain $\hat{\Omega}_{k+1}$ by solely using the bounds L_{k+1} and U_{k+1} , we can store the narrow band efficiently without having to use the original domain. To this end we make use of a self referential array of the following structs:

```
struct { float value; int vox_index, n1, n2, n3; } NB;
```

We calculate the indices of the local neighbourhood of each voxel according to the narrow band and store these in the variables `n1`, `n2` and `n3`, which allows us to calculate the differential operators in algorithm 6. The variable `vox_index` stores the index that the struct element would have in the original voxel space. This is necessary for the transformation in step 1. In contrast to a linked list implementation the array implementation makes the approach easily parallelizable, since we can pass each NB `struct` element independently to different cores.

4 Experimental Evaluation

In this section we perform experimental evaluations of the narrow band method. To this end we will provide qualitative and quantitative results in order to empirically show that using a small bandwidth i.e. a fraction of the original domain we can approximate the true minimizer of the energy to an extent that most of the times is neither visible nor quantitative differences are to be seen. Additionally, we show that using merely a consumer low end card we are able to reconstruct depth information with a resolution that is not even possible on high end GPU's. The approach also scales well with recent graphics cards. Lastly we elaborate on the loss of fine details depending on the NBT. (c.f. Figure 1).

4.1 Setup

The hardware setting for our experiments is a Intel(R) Core(TM)2 Quad CPU Q8300 @ 2.50Ghz computer equipped with a low end graphics card of the type NVIDIA GeForce GTX 285 with 1 GB memory, which possesses 240 CUDA cores. The parallel implementation of the narrow band algorithm is based on NVIDIA's CUDA framework. We set the parameter for the dataterm to the value $\lambda = 50$ for all *TV* computations. All the results are based on a 2 scale *NB* method which works well for both small and large input stereo pairs. For the evaluation we used image pairs from the Middlebury datasets [16].

4.2 Memory Efficiency

Based on our implementation, we provide a memory demand estimator (*MDE*), for both the dense as well as the *NB* approach:

$$\begin{aligned} MDE_{dense} &= \text{height} \cdot \text{width} \cdot (\text{levels} \cdot 24\text{byte} + 28\text{byte}) \\ MDE_{NB} &= \text{height} \cdot \text{width} \cdot (\bar{C} \cdot 56\text{byte} + 36\text{byte}) \end{aligned}$$

The factor \bar{C} is the average number of voxels to be stored for each pixel in order to encode the narrow band around the graph of indicator variable u in our application. The graph is a manifold that describes the surface of the reconstructed 3D scenery which is bigger than the image size itself. Because of that \bar{C} is at least as large as the chosen *NBT*. Experimentally \bar{C} is in the order of the *NBT*. For example, the values for \bar{C} in the experiments in Figure 4 are



Fig. 3. Left input image¹ and solution for $NBT = 4$, with 500 iterations. Computation time is 31.7 s.

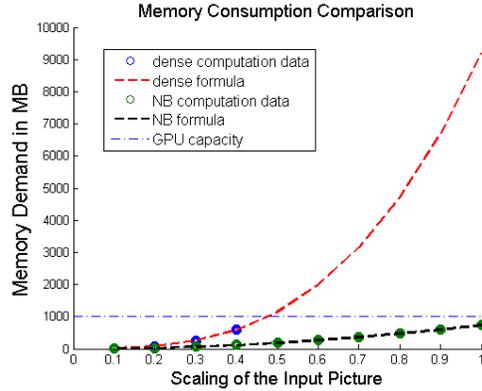


Fig. 4. The plot shows the memory consumption for different problem sizes. Scaling=0.1 stands for 152 x 132 pixels and 20 levels and scaling=1 corresponds to 1520 x 1320 x 200. Figure 3 shows the left input image and a solution example.

in the range 5.8 to 6.3 for an NBT of 4. Note that in the NB approach the third factor only depends on \bar{C} . Hence we practically obtain a complexity of a variational problem in 2D. The above estimators are specific for our algorithmic approach, but the core idea that the complexity is independent of the number of levels is the same with any implementation.

4.3 Runtime Evaluation

For comparison we take a dense reconstruction (i.e. on the full domain) obtained after 15000 iterations as reference. We use 80 values for the discretization of the disparity space Γ . Figure 5 shows the left input image and a reconstruction example. To measure the quality of a solution we use the average pixel color difference between the reference result and the computed disparity. Figure 6 shows the runtime convergence comparison between the NB approach and the computation on a dense label space. We observe that the NB method outperforms the dense approach in terms of runtime.

4.4 Narrow Band Thickness and Fine Details

In this section we provide an empirical correlation between the NBT and the quality of the reconstruction. Figure 7 (b) shows the quality of reconstructions using NBT 's between 4 and 12 depending on the number of iterations. The plot shows that the quality increases with the NBT . Problems arise for fine detail structures and a low NBT as can be seen in the close up pictures of the reindeer for different NBT 's in Figure 7. Only at a NBT of 10 the area around the legs of the reindeer is well represented. Therefore the approach introduces the drawbacks of classical coarse to fine schemes to multi-label optimization. This issue remains a challenging future work.



Fig. 5. The Solution corresponds to the black circle in Fig. 6.

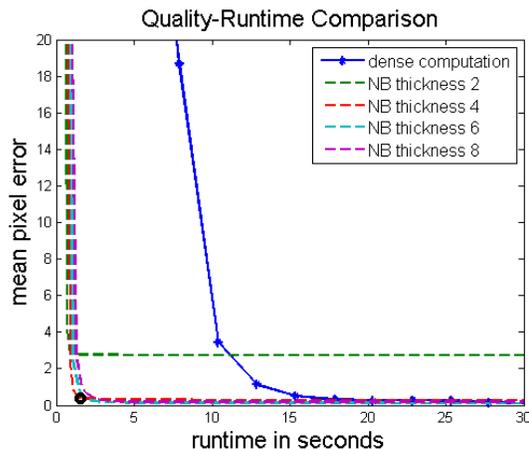


Fig. 6. Quality vs runtime plots for an 444×372 input image using 80 disparity levels, see Figure 5. The comparison is done by using pixel mean value error.

5 Conclusion

We provided a systematic experimental validation of narrow band methods for variational multilabel optimization. We revisited convex relaxation approaches to variational multilabel optimization, presented a coarse-to-fine narrow band formulation and experimentally evaluated the tradeoff between accuracy of recovered solutions on one hand and speed and memory requirements on the other. Considering the simplicity of our functional, our experiments demonstrate that the narrow band reformulation allows to reduce memory and computation time by orders of magnitude with moderate loss in accuracy. These experiments indicate that in conjunction with narrow band formulations, convex relaxation techniques for multilabel optimization exhibit the potential for solving large scale reconstruction problems.

References

1. Adalsteinsson, D., Sethian, J.A.: A fast level set method for propagating interfaces. *Journal of Computational Physics* 118, 269–277 (1994)
2. Baeza, A., Caselles, V., Gargallo, P., Papadakis, N.: A narrow band method for the convex formulation of discrete multilabel problems. *Multiscale Modeling & Simulation* 8(5), 2048–2078 (2010)
3. Barnard, S.: Stochastic stereo matching over scale. *International Journal of Computer Vision* 3(1), 17–32 (1989), <http://dx.doi.org/10.1007/BF00054836>
4. Chambolle, A., Cremers, D., Pock, T.: A convex approach to minimal partitions. *J. Imaging Sci.* 5(4), 1113–1158 (2012)
5. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision* 40(1), 120–145 (2011)

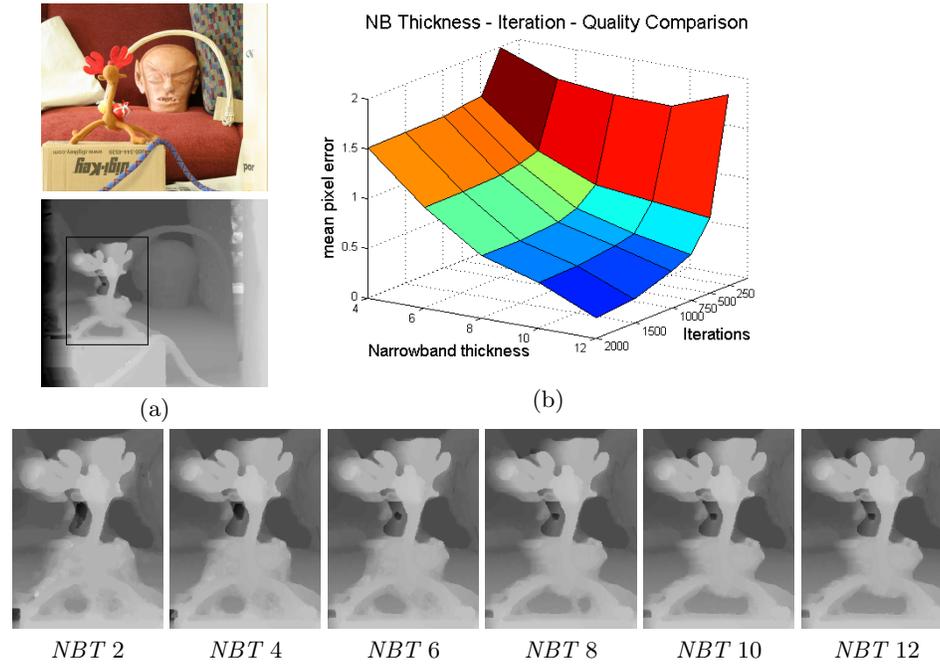


Fig. 7. (a) A 448×372 input image and the reconstruction on a dense 80 label space. (b) quality comparison for different *NBT*'s. Below are 6 close up pictures of the reindeer for different *NBT*'s.

6. Goldluecke, B., Strelakovsky, E., Cremers, D.: Tight convex relaxations for vector-valued labeling. *SIAM Journal on Imaging Sciences* (2013), to appear
7. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30(2), 328–341 (2008)
8. Horn, B., Schunck, B.: Determining optical flow. *A.I.* 17, 185–203 (1981)
9. Ishikawa, H.: Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(10), 1333–1336 (2003)
10. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
11. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. *J. Imaging Sci.* 4(4), 1049–1096 (2011)
12. Lombaert, H., Sun, Y., Grady, L., Xu, C.: A multilevel banded graph cuts method for fast image segmentation. In: *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 - Volume 01*. pp. 259–265. ICCV '05, IEEE Computer Society, Washington, DC, USA (2005)
13. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: *ICCV* (2011)
14. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. *SIAM J. Imaging Sciences* 3(4), 1122–1145 (2010)
15. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: *Proceedings of the 10th European*

- Conference on Computer Vision: Part III. pp. 792–805. ECCV '08, Springer-Verlag, Berlin, Heidelberg (2008)
16. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: CVPR (2007)
 17. Zach, C., Gallup, D., Frahm, J.M., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: Workshop on Vision, Modeling and Visualization (October 2008)