

Superpixels, Occlusion and Stereo

Yuhang Zhang
Australian National University
Yuhang.Zhang@anu.edu.au

Richard Hartley
Australian National University
and National ICT Australia *

John Mashford, Stewart Burn
CSIRO, Melbourne
John.Mashford@csiro.au
Stewart.Burn@csiro.au

Abstract

Graph-based energy minimization is now the state of the art in stereo matching methods. In spite of its outstanding performance, little effort has been made to enhance its capability of occlusion handling. We propose an occlusion constraint, an iterative optimization strategy and a mechanism that proceeds on both the digital pixel level and the superpixel level. Our method explicitly handles occlusion in the framework of graph-based energy minimization. It is fast and outperforms previous methods especially in the matching accuracy of boundary areas.

1. Introduction

We propose a new method for binocular stereo matching. Our method is based on graph-based energy minimization algorithms, which were popularized by Boykov *et al.* [3] and are now considered as the state of the art among all methods of stereo matching [15]. We incorporate superpixels into stereo matching, propose a new constraint for occlusion and explicitly handle occlusion in the framework of energy minimization. Our method gives promising results on the benchmark dataset [15].

Occlusion has long been regarded as one of the most challenging issues in stereo matching. Since occlusion usually appears together with depth discontinuity, properly handling occlusion and accurate boundary detection are usually two coupled problems. However, most graph-based methods [3, 4, 5, 8, 9] generally overlook occlusion in spite of their strong performance. Considering the frequent existence of occlusion in almost all non-trivial stereo pairs, especially in the case of wide base-line stereo, developing a method which can properly handle it in the framework of graph-based energy minimization is of high value. Note that the graph-based methods generally model stereo matching

as a multi-labeling problem. One solution proposed in the previous approaches to occlusion is to create an extra label for the occluded pixels [6]. However, as this extra label has completely different meaning compared to the other disparity related labels, manipulating them together in the same energy function cannot guarantee a reliable result, which has been evidenced by Kang *et al.* [6] as well. A more sound solution proposed by Kolmogorov and Zabih [7] is to use a completely new construction of the energy function, in which each vertex in the graph corresponds to a labeling rather than a pixel. Occlusion is no longer a singular case in that construction. However, as most relevant work is based on the old construction, this new construction makes it harder to maintain their advantages, such as a robust smoothing term. Before the emergence of graph-based methods, several different mechanisms had been used for occlusion detection. The order constraint [1] is powerful, but only holds for particular images. The cooperative method [21] detects occlusion through thresholding the matching strength, which is risky as the false matching generated by the occluded points can be strong as well. A more classic way to handle occlusion is cross-checking [2], which explicitly enforces the requirement of uniqueness. We combine this classic mechanism with our new proposed occlusion constraint. Occlusion can then be detected during iterative energy minimization following the old energy function construction. This new mechanism for occlusion handling is our major contribution. In the output of our method, occluded pixels will not only be explicitly marked, but also be assigned probable disparities as well.

Superpixel partitioning, or more generally speaking, segmentation is not new to stereo matching as a preprocessing step. Under the assumption that pixels in a homogeneous region should have similar and continuous disparities, pre-segmentation can largely reduce the sequential computation load, reduce the ambiguity during pixel matching, especially in the areas of weak textures, and consolidate object boundaries, which are vulnerable to the smooth prior. Several previous works on stereo matching followed this philosophy [18, 22]. However, segmentation has not been in-

*NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

corporated into the graph-based methods so far. In order to avoid unnecessary flattening, over-segmentation is usually preferable to under-segmentation. In particular, superpixels [13] split the original image into a large number of small patches of regular shapes and comparable sizes. Over the years, more and more faster implementations [10, 11, 19] have been proposed making superpixels practical for more and more applications. In our method, we use superpixels to boost the processing speed as well as to consolidate object boundaries.

2. Algorithm Outline

1. Initialize an occlusion map for each of the two images in the stereo pair respectively, with all pixels marked as zero, meaning not occluded.
2. Over-segment each of the two images into superpixels.
3. Generate the unary and binary terms for the superpixels in the two images according to the occlusion map.
4. Implement superpixel-level stereo matching through graph-based minimization, using each of the two images as reference respectively.
5. Implement digital-pixel-level cross-checking to update the occlusion map.
6. Loop Steps 3, 4 and 5 until the occlusion map becomes stable.

A more intuitive illustration of the general structure of our method is given in Figure 1, which shows the transformations between different data. The numbers attached to each transformation correspond to steps in the outline above. Several new features can be found in the outline of our algorithm. Firstly, we treat the two images in the stereo pair symmetrically. Secondly, we work on the superpixels and the digital pixels alternatively. Lastly, as the algorithm proceeds, it modifies the energy function according to the information it acquires and minimizes the function multiple times. The intention behind this design will be explained in the subsequent sections. In the following parts of this paper, we first review the general framework of the graph-based methods, analyze its limitation, especially how occlusion is treated, and then explain our new method in details. Experiments and conclusions will then be presented.

3. Graph-based Energy Minimization

To model a stereo matching problem using energy minimization, one first selects a reference image from the image pair. Each pixel in the reference image is then modeled as a variable in a Markov Random Field (MRF), whose value or label corresponds to its disparity. The unary cost (data term)

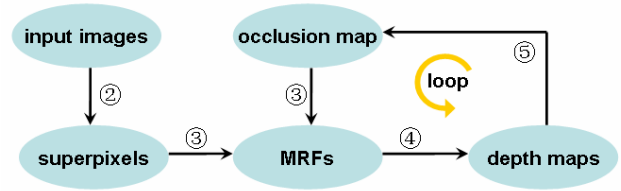


Figure 1. A diagram showing the transformation between different data in the proposed method. The number attached to each transformation indicates its corresponding step in the algorithm outline.

D_i of this variable is usually determined by the colour difference between the two pixels matched with by the current disparity value. The binary cost (smoothing term) S_{ij} will be added as well to impose the smooth prior over neighbour pixels, *i.e.* causing them tend to have similar value. The disparity of all pixels in the reference image will then be determined through globally minimizing the energy within the MRF, as show in (1), where N is the set of all neighbours in the MRF, λ is a positive constant balancing the weight of the unary terms and the binary terms.

$$E = \sum_i D_i + \lambda \sum_{i,j \in N} S_{ij} \quad (1)$$

One notable limitation with the above model is that it treats the stereo pair asymmetrically. Recall that two important constraints were imposed to stereo matching as early as the work of Poggio *et al.* [12], *i.e.* the uniqueness constraint and the continuity constraint. The uniqueness constraint states that each pixel in the two images should have a unique disparity. The continuity constraint requires the disparities of neighbour pixels to be continuous almost everywhere. Whereas the above model enforces the continuity constraint through the smoothing terms, it leaves the uniqueness constraint unchecked during its single-way matching. In particular, when each pixel in the reference image is assigned a unique disparity, no mechanism prevents them from being associated with the same pixel in the other image. Hence, a pixel in the other image might have multiple matches in the reference image and therefore have more than one disparity. Moreover, as the matching is implemented along one direction only, occlusion can hardly be detected. All occluded pixels are treated as visible pixels, forced to find correspondences that do not really exist in the other image. A desirable outcome is that the true matches surrounding a given pixel can somehow help them find their correct disparities, even though the matches cannot be true. However, things can easily turn out to be the opposite, especially when a large area of occlusion is present. These occluded pixels find false matches and further mislead the other pixels nearby through the smoothing terms.

Therefore, we propose to treat the two images in a stereo pair symmetrically and implement the stereo matching in both directions. However, graph-based algorithms are not by nature fast. That is the basic reason why we use superpixels.

4. Superpixels for Stereo Matching

As indicated by the algorithm outline, we need to implement graph-based energy minimization in both directions for multiple iterations. Therefore, it is necessary to find a way to implement each single minimization in short time. The hierarchical MRF optimization algorithm proposed by Zhang *et al.* [20] is fast but usually distorts small structures due to evenly merging neighbour pixels. Instead, we choose to merge selective neighbour pixels in a more meaningful way, *i.e.* through superpixels. The graph-based method [19] models superpixel cropping as a local small-range labeling problem, which can be solved much more efficiently than the original N-Cuts [17] based method and produce comparable results. Some segmentation examples are shown in Figure 2. We incorporate it into our method. In our exper-



Figure 2. Two segmentation examples: the left is the original image and the right is the resulted superpixels.

iment, segmenting a 695×555 images into around 4,000 superpixels takes about 10 seconds on a machine equipped with a single core 1.6G AMD CPU and 1G RAM. This speed is acceptable as we only need to implement it once for each image in a stereo pair. Later we will see, that it is the main part of the processing time of our method. A larger number of superpixels can improve the matching accuracy at the cost of a lower efficiency.

Another reason for using superpixels is to consolidate the boundary at a depth discontinuity. Although the continuity constraint is only applicable almost everywhere, what people really impose is a smoothing term everywhere. Un-

wanted fattening and shrinkage is therefore widely seen in graph-based stereo matching (see Figure 3). To reduce this



Figure 3. The drifting of the depth boundaries fattens and shrinks areas of homogeneous depth. The right image is a depth map produced by α -expansion. The brush next to the jar becomes thicker, whereas the sticks in the cup partly disappear. Blurry boundaries are quite common.

side-effect, previous works generally appeal to the idea of edge-preservation. In our method, superpixels semantically separate pixels near colour boundaries into homogeneous groups. As depth boundaries usually coexist with colour boundaries (otherwise we cannot detect a depth discontinuity), the grouping due to superpixels can confine the drift of depth boundaries. Superpixels can also remove the small depth differences within a homogeneous region. However, we will not worry about it too much, as Zhang *et al.* has already shown that details of this size can be retrieved by cheap local fine tuning [20].

In the next section we discuss the data term and the smoothing term for a superpixel. We deliberately skip the content of occlusion map for the moment, because its generation and usage will become more intuitive in Step 5.

5. Data Term & Smoothing Term

We now discuss the construction of the energy function for the superpixels. The data term \mathcal{D} for a superpixel I having label L is simply the aggregation of the data terms for each of its member digital pixels having label L , as shown in (2).

$$\mathcal{D}_I(L) = \sum_{i \in I} D_i(L) \quad (2)$$

However, the smoothing term \mathcal{S} for superpixels is a bit more complex. Whereas the strength of the edge between each pair of digital pixels is usually constant or only affected by their colour difference, the edge strength between each pair of superpixels is also affected by the length of the boundary between them. We assume the digital pixels in an image are 4-connected. The connection strength \hat{s} between a digital pixel i and a superpixel J is measured by the number of digital pixels which are the members of J but exist in the neighbourhood of i , as shown in (3), where \mathcal{M} and \mathcal{N} are both boolean-valued functions judging whether the two

input pixels are members or neighbours of each other.

$$\hat{s}(i, J) = \sum_{\forall j, \mathcal{N}(i, j)=1} \mathcal{M}(j, J), \quad (3)$$

$$\mathcal{M}(j, J) = \begin{cases} 1 & \text{if } j \in J \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$\mathcal{N}(i, j) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Then the connection strength \hat{S} between two superpixels can be defined as

$$\hat{S}(I, J) = \sum_{i \in I} \hat{s}(i, J). \quad (6)$$

It can be easily verified that $\hat{S}(I, J) = \hat{S}(J, I)$. Further denote the average colour, after normalization, of the digital pixels in superpixel I as $\mathcal{C}(I)$. Our smoothing term \mathcal{S} for superpixel I and J with label L_I and L_J respectively is defined as

$$\mathcal{S}_{ij}(L_I, L_J) = \mathcal{S}(I, J) |L_I - L_J| (1 - |\mathcal{C}_I - \mathcal{C}_J|) \quad (7)$$

According to (7), the longer the boundary between the two superpixels is, the stronger their smoothing term is; the more similar their average colour is, the stronger their smoothing term is. Furthermore, to preserve sharp edges, we can modify (7) into a truncated smoothing term by setting an upper boundary. The label range for superpixels is the same as that for digital pixels.

6. Energy Minimization

After the generation of superpixels, the number of variables in the MRF generally decreases by two orders of magnitude. The number of edges, due to the regular shapes of the superpixels, decreases by about the same ratio as well. Whereas optimizing an MRF of size 1000×1000 takes some time, optimizing an MRF of size smaller than 100×100 will be much faster. It can be expected that all energy minimization algorithms proposed in recent years can converge quickly. In our implementation, α -expansion takes approximately 2 seconds to complete a single minimization. Therefore, implementing bi-direction stereo matching for 3 iterations only takes about 12 seconds.

7. Handling Occlusion

We now explain the core part of our method, the detection and the disposal of occlusions. After the stereo matching in both directions, we obtain depth maps \mathcal{P}_l and \mathcal{P}_r for pixels in both of the images. The two depth maps can be inconsistent at some of the pixels. We use cross-checking to

detect them and record them with occlusion maps according to (8) and (9), where m is a tolerance parameter setting the maximum allowed error between a pair of consistent matches.

$$\mathcal{O}_l(u, v) = \begin{cases} 0 & \text{if } |\mathcal{P}_l(u, v) - \mathcal{P}_r(u - \mathcal{P}_l(u, v), v)| \leq m \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

$$\mathcal{O}_r(u, v) = \begin{cases} 0 & \text{if } |\mathcal{P}_r(u, v) - \mathcal{P}_l(u + \mathcal{P}_r(u, v), v)| \leq m \\ 1 & \text{otherwise} \end{cases} \quad (9)$$

Usually, m is usually set to zero, namely no error is allowed for a pair of matches to be consistent. However, in our case, as stereo matching is implemented at the superpixel level, and that it is unavoidable for superpixels to merge pixels with small depth difference together, zero tolerance will cause many pixels to be unable to find a consistent match. We therefore set the value of m to one. This minor difference will lead to significant improvement during experiment.

In an occlusion map, the pixels with value zero suggest that the two disparity maps are consistent; they are visible in both images, and their current disparities are highly reliable. However, pixels with value one have diverse interpretations. They might be occluded pixels, which have no correspondences in the other image, or simply visible pixels with incorrect disparities. We use another iteration of energy minimization to separate them from each other. To achieve this target, we modify the energy function based on the information we collect from the last run of energy minimization. But first of all, we shall make the occlusion constraint clear.

7.1. The Occlusion Constraint

Whereas occlusion involves at least two pixels, *i.e.* the occluded one and the occluding one, previous work only paid attention to the occluded ones. That is, they explicitly identify those occluded pixels without caring what pixels occlude them. A possible situation might be that a pixel is labeled as occluded, however, according to the disparities of the visible pixels no one could possibly occlude it (see Figure 4). We hence propose an occlusion constraint,



Figure 4. Five horizontally continuous pixels on the right image are labeled as above. The disparity range in the problem is between 0 and 1. Whereas two pixels are labeled as occluded, only the left one can be occluded.

that within the image boundaries each occluded pixel must be occluded by a visible pixel at a smaller depth. Like the uniqueness and continuity constraints, our occlusion constraint only states a very obvious fact. However, it will

show impressive power during stereo matching. We next show how to modify the energy function based on it.

7.2. Energy Function Modification

As claimed by the occlusion constraint, each occluded pixel must be occluded by a visible pixel. If the disparity range is from 0 to $n - 1$, for each pixel there are $n - 1$ pixels that have the chance to occlude it. Therefore to impose the occlusion constraint without knowing in advance which pixel will do the occluding we need to add an n -th order term which uses the labels for all the n pixels in the energy function. Obviously that will make the energy function too complex to minimize. Instead, we appeal to the result of the previous energy minimization iteration. Pixels having 0 value in the occlusion map are those that are visible in both images with reliable disparities. Assume the pixel at (u, v) in the right image has value 1 in the occlusion map \mathcal{O}_r . Its potential labels can be divided into three groups according to the criteria below:

$$\begin{aligned} G_1 &= \{L | \mathcal{O}_l(u + L, v) = 0, \mathcal{P}_l(u + L, v) > L\} \quad (10) \\ G_2 &= \{L | \mathcal{O}_l(u + L, v) = 0, \mathcal{P}_l(u + L, v) < L\} \quad (11) \\ G_3 &= \{L | \mathcal{O}_l(u + L, v) = 1\}. \quad (12) \end{aligned}$$

In (10), $\mathcal{O}_l(u + L, v) = 0$ suggests that the label L associates pixel (u, v) with a pixel that already has a reliable match. Moreover, $\mathcal{P}_l(u + L, v) > L$ suggests this reliable match has a larger disparity, namely a smaller depth, than pixel (u, v) . Therefore, pixel (u, v) is occluded in the left image according to the occlusion constraint. Differently, (11) suggests pixel (u, v) occludes a visible pixel, which is absolutely an improper label. The last group suggests pixel (u, v) is associated to another pixel that has no reliable match so far. That really means pixel (u, v) is not occluded in the left image, otherwise it should not have a match. Accordingly, we modify the unary cost of the three groups of labels as follows:

$$D'_i(L) = \begin{cases} c & \text{if } L \in G_1 \\ +\infty & \text{if } L \in G_2 \\ D_i(L) & \text{if } L \in G_3 \end{cases} \quad (13)$$

where

$$c = \min_{L \in G_3} D_i(L). \quad (14)$$

According to (13), pixels currently having no reliable labels will not occlude pixels currently having reliable labels. It costs the same to be occluded or to have the best match among unsettled pixels in the other image. Recall our method uses superpixels. If a superpixel is only partially occluded, the non-occluded part will help the occluded part make the decision. If a superpixel is completely occluded, its smoothing term with the neighbouring superpixels will help it make the decision. Therefore, whether a single pixel

is occluded is determined by the optimal state of the MRF. We find our way of finding occlusion more natural than the usage of an extra label [6] and the new construction of the energy function [7]. While using an extra label [6], a heuristic and constant cost is applied to all pixels having the occlusion label. In the new construction of the energy function [7], the unary cost for being occluded is still the colour difference between the two pixels related by the label, which does not make much sense.

8. Experiment

To demonstrate the capability of our method on stereo matching, especially with occlusion, we select a set of stereo pairs from the Middlebury stereo dataset [14, 16]. The selected images all have strong depth discontinuity and occlusion. The minimum disparity range in the selected images is 60. On these images the original graph-based methods generally make the mistake of unnecessary fattening and shrinkage as shown in Figure 3.

8.1. Superpixel-based Stereo

Let us now see what can superpixels do first. In Figure 5, we implement α -expansion on both the digital pixel level as well as on the superpixel level. The same parameters are used for the two methods when applicable.

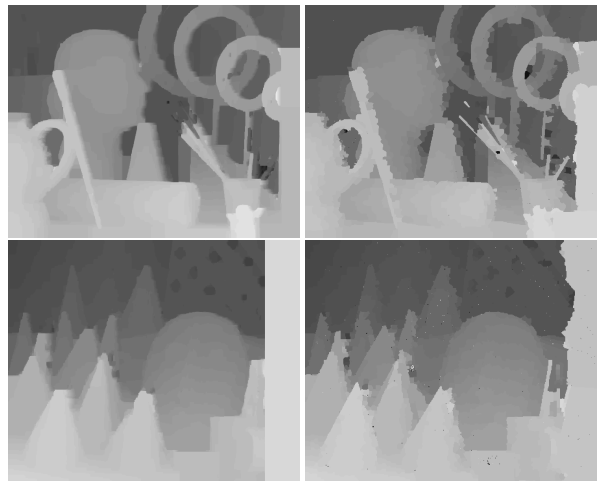


Figure 5. Left: the depth map produced by the original α -expansion. Right: the depth map produced by superpixel level α -expansion; no occlusion constraint is applied.

The quality of the results of the two methods are in general comparable. Compared to the digital pixel level α -expansion, the superpixel level α -expansion protects small structures like sticks and cone tops better but sometimes also makes errors near depth boundaries more distinct. The blurry boundaries generated by the digital pixel level α -expansion actually contains errors as well, but can easily

escape attention. One may propose the use of a smoothing term other than the linear one we use to improve the boundary accuracy. However, no matter whether it is a convex or concave smoothing term, they both have their own bias towards a particular structure and destroy the other types. For example, a concave term will encourage sharp jumps but destroy gradual changes and a convex term will do the opposite. A linear smoothing term appears to be more non-aligned.

Despite the comparable quality, the superpixel level α -expansion is significantly faster than the digital pixel level α -expansion. On the cone image, the running times of the two methods are 40 seconds and 24 seconds respectively. On the art image, the running times of the two methods are 88 seconds and 29 seconds respectively. The time of superpixel segmentation is counted in. The art image takes longer because it is larger. However, an interesting observation is that the time taken by the superpixel level α -expansion does not increase much. That is because most of the time is spent on superpixel segmentation but not on stereo matching. The running time of graph-based superpixel cropping is less sensitive to the image size than the graph-based stereo matching. This fact will benefit us during iterative stereo matching.

8.2. Iterative Stereo

Now we show how our iterative method performs on occlusion. We first set the m in (8) and (9) to 0. Still, using the art and the cone image as the examples, Figure 6 shows the evolution of the depth map and the occlusion map over the first 3 iterations. The number under each image indicates what percentage of the digital pixels have reliable matches according to the current depth estimation.

Significant progress can be observed between the first and the second iteration. That is, when the first batch of occlusions were detected, a large number of visible pixels, previously misled by those occlusions, find their true matches. As more visible pixels find their reliable matches, more occlusion can be detected. Such a chain reaction goes on. Smaller progress can still be made even after the 5th iteration. By visually checking, we notice that after the second iteration, almost all fattened regions recover to their actual shape, e.g. the conical flask in the art image, and the cones in the cone image. Most of the pixels besides these occluding objects which are to be occluded in the other image have been assigned proper disparity as well. We also notice that our iterative method improves the matching accuracy on the art image particularly much. That is because the art image is more complex than the cone image, which is full of depth discontinuity and occlusion.

Some pixels which should not be occluded remain white in the occlusion map. This is because of merging pixels with different depth into the same superpixel, and the usage

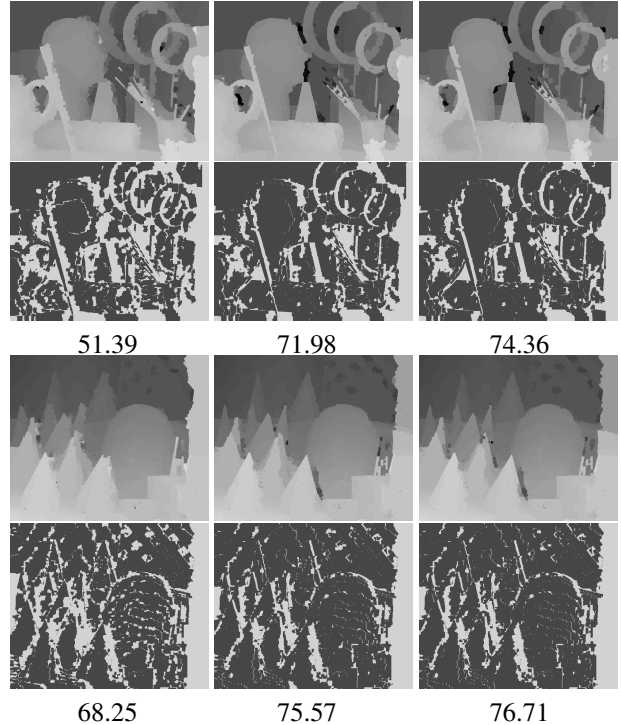


Figure 6. Evolution of the depth map (first row) and the occlusion map(second row) during the first three iterations of the proposed method. The number below each image records the percentage of reliable matches (dark points) in the image above. Results for $m = 0$.

of a zero tolerance parameter m . The inaccuracy of superpixel segmentation can contribute to these mistakes as well. Future work might improve the accuracy of superpixels.

We now change the value of m from 0 to 1 and repeat the same experiment. Figure 7 shows the new results. The percentage of reliable matches is increased. This is not surprising since we are using a larger tolerance parameter. What really improves is the accuracy of the occlusion map, as well as the boundary area in the depth map. As shown by the image, after the 3rd iteration almost all the objects now have clear and correct boundaries. Moreover, more pixels which should be occluded by these objects are assigned correct disparity as well. The white points remaining in the occlusion map almost all correspond to occluded pixels. The reason why such an improvement can be made has already been analyzed in Section 7. More results produced by our method can be found in Figure 8 and 9.

9. Conclusion

Earlier in this paper, we pointed out the three novel aspects of our method compared to other graph-based methods. The intention behind them should be clear by now. We treat the two input images symmetrically, so that oc-

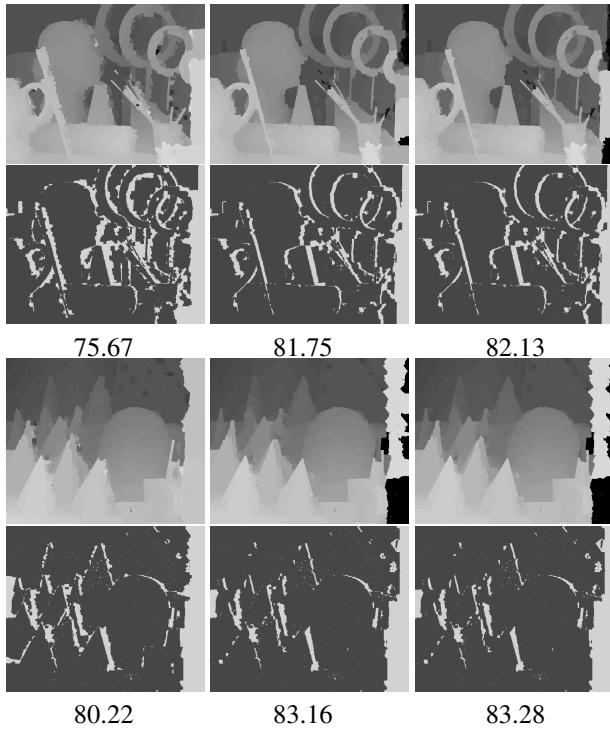


Figure 7. The evolution of the depth map (first row) and the occlusion map (second row) during the first three iterations of the proposed method. The number below each image records the percentage of reliable matches (dark point) in the image above. $m = 1$.

clusion and false matches can be detected. We process superpixels and digital pixels alternatively, so that efficiency and the accuracy can be maintained at the same time. We make the energy function modifiable so that the occlusion constraint can be embedded into energy minimization without increasing its complexity. Furthermore, whereas there is always an ambiguity between occlusion and false matching of visible pixels, our iterative minimization focuses on the reliable matches to distinguish these cases. We believe that the modifiable energy function and the alternative processing of digital pixels and superpixels will be valuable in future work as two general strategies in graph-based stereo matching. The later is a variant of the classic hierarchical mechanism.

We see future work in a more close collaboration between superpixel and stereo matching. During the maturation of the depth map, there is a chance to revise the configuration of the superpixels through further division and combining. Improvement to the superpixel segmentation will then reward the stereo matching, forming another chain reaction. Higher accuracy on both sides can then be expected.

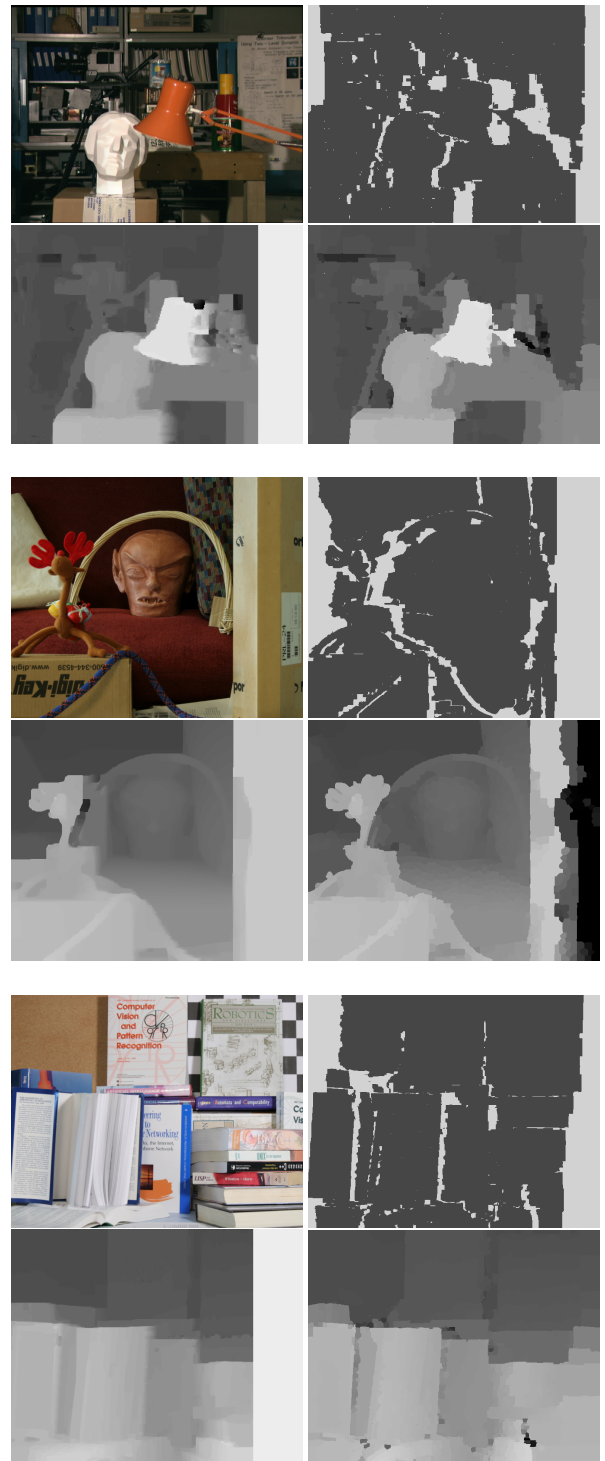


Figure 8. Each group in clockwise order: right image; occlusion map after the 3rd iteration; depth map after the 3rd iteration; depth map produced by the original α -expansion.

References

- [1] P. Belhumeur and D. Mumford. A bayesian treatment of the stereo correspondence problem using half-occluded regions.

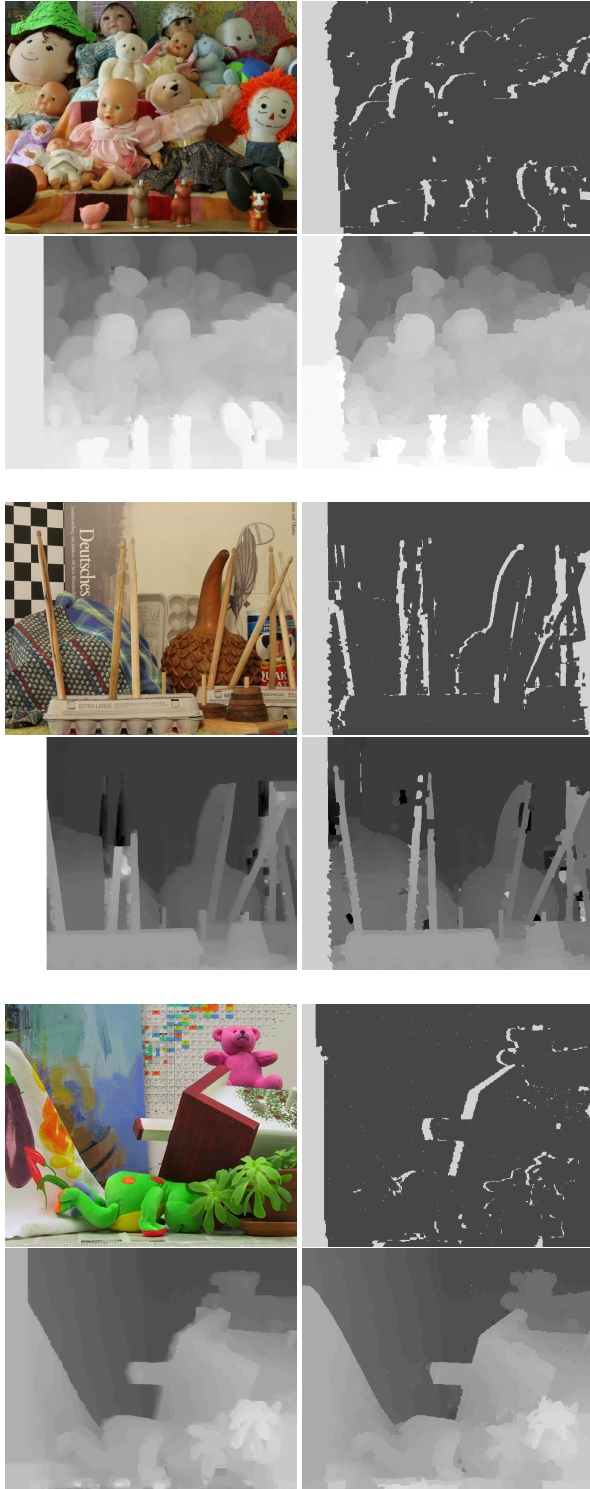


Figure 9. Each group in clockwise order: left image; occlusion map after the 3rd iteration; depth map after the 3rd iteration; depth map produced by the original α -expansion.

- In *CVPR*, pages 506–512, 1992. 1
- [2] R. C. Bolles and J. Woodfill. Spatiotemporal consistency checking of passive range data. In *International Symposium on Robotics Research*, 1993. 1
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001. 1
- [4] P. Carr and R. Hartley. Minimizing energy functions on 4-connected lattices using elimination. In *ICCV*, 2009. 1
- [5] P. Carr and R. Hartley. Solving multilabel graph cut problems with multilabel swap. In *DICTA*, 2009. 1
- [6] S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *CVPR 2001*, 2001. 1, 5
- [7] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, pages 508–515, 2001. 1, 5
- [8] N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(8):1436–1453, 2007. 1
- [9] N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. *Comput. Vis. Image Underst.*, 112(1):14–29, 2008. 1
- [10] A. Levinstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(12):2290–2297, 2009. 2
- [11] A. P. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones. Superpixel lattices. In *CVPR*, 2008. 2
- [12] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–319, 1985. 2
- [13] X. Ren and J. Malik. Learning a classification model for segmentation. In *ICCV '03*, page 10, 2003. 2
- [14] D. Scharstein and C. Pal. Learning conditional random fields for stereo. *CVPR*, 0:1–8, 2007. 5
- [15] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47(1-3):7–42, 2002. 1
- [16] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. *CVPR*, 1:195, 2003. 5
- [17] J. Shi and J. Malik. Normalized cuts and image segmentation. In *CVPR '97*, page 731, 1997. 3
- [18] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. *IEEE International Conference on Computer Vision*, 1:532, 2001. 1
- [19] O. Veksler, Y. Boykov, and P. Mehrani. Superpixels and supervoxels in an energy optimization framework. In *ECCV 2010*, volume 6315, pages 211–224. 2010. 2, 3
- [20] Y. Zhang, R. Hartley, and L. Wang. Fast multi-labelling for stereo matching. In *ECCV 2010*, pages 524–537. 2010. 3
- [21] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:675–684, July 2000. 1
- [22] C. L. Zitnick and S. B. Kang. Stereo for image-based rendering using image over-segmentation. *Int. J. Comput. Vision*, 75(1):49–65, 2007. 1