Global Stereo Reconstruction under Second Order Smoothness Priors

O. J. Woodford[†] P. H. S. Torr[‡] I. D. Reid[†] A. W. Fitzgibbon[⋆]

[†]Department of Engineering Science, University of Oxford
{ojw,ian}@robots.ox.ac.uk

[‡]Department of Computing, Oxford Brookes University
philiptorr@brookes.ac.uk

[⋆]Microsoft Research, Cambridge, U.K.
awf@microsoft.com

Abstract

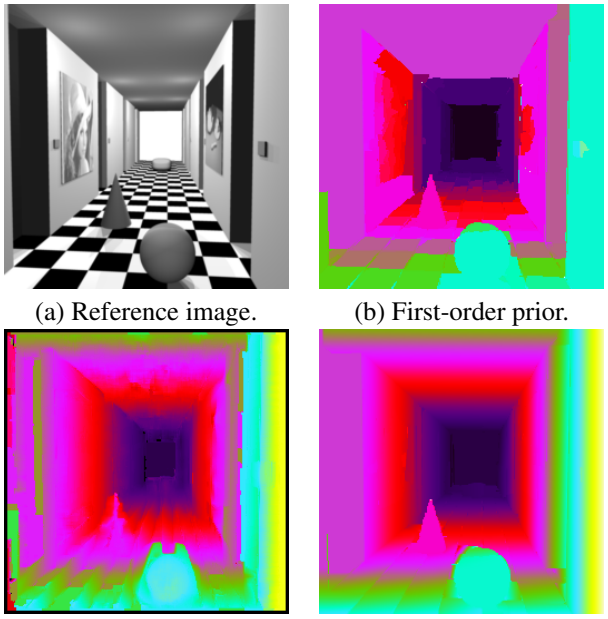
Second-order priors on the smoothness of 3D surfaces are a better model of typical scenes than first-order priors. However, stereo reconstruction using global inference algorithms, such as graph-cuts, has not been able to incorporate second-order priors because the triple cliques needed to express them yield intractable (non-submodular) optimization problems.

This paper shows that inference with triple cliques can be effectively optimized. Our optimization strategy is a development of recent extensions to α-expansion, based on the "QPBO" algorithm [5, 14, 26]. The strategy is to repeatedly merge proposal *depth maps using a novel extension of QPBO. Proposal depth maps can come from any source, for example fronto-parallel planes as in α-expansion, or indeed any existing stereo algorithm, with arbitrary parameter settings.*

Experimental results demonstrate the usefulness of the second-order prior and the efficacy of our optimization framework. An implementation of our stereo framework is available online [34].

1. Introduction

Multiple-view dense stereo has made considerable progress in recent years, in part because the problem can be cast in an energy minimization framework for which there exist inference algorithms that can efficiently find good (if not always global) minima. Algorithms based on graph cuts, in particular, can incorporate *visibility reasoning* as well as *smoothness priors* into the estimation of depth maps. However, the smoothness priors used in graph-cut based estimates have to date been first-order priors, which favor low-curvature fronto-parallel surfaces—indeed, the prior is maximized by fronto-parallel planes. Even in man-made scenes, this is far from accurate, as illustrated in figure 1, and leads to inaccurate depth estimates. It has long been known [3, 13, 30] that a second order smoothness prior can better model the real world, but it has not yet been possible to combine visibility reasoning and second-order smoothness in an optimization framework which finds good optima.



(a) Reference image. (b) First-order prior.

(c) Li & Zucker's result. (d) Our result.

Figure 1. **Second-order smoothness priors**. (a) A reference image for which we wish to produce a dense depth map. (b)–(d) The disparity (inverse depth) maps produced by (b) first-order prior, (c) second-order prior of Li and Zucker [23] and (d) second-order prior with visibility, optimized as in this paper.

The main contribution of this paper is to introduce an effective optimization strategy for stereo reconstruction with triple cliques. This means that visibility reasoning and second-order priors can be combined for the first time. We show that this algorithm produces excellent results both on the Middlebury test set [27] and on real-world examples with curved surfaces.

1.1. Background

Second order smoothness priors for stereo reconstruction have a long history. Grimson [13] and Terzopoulos [30] both proposed second order priors for stereo in the early 1980s, in the form of the thin plate model. This was extended to the piecewise second order "weak plate" model by Blake and Zisserman [3], and recently Ishikawa and

Geiger [16] have argued that second order priors may be closer to those that the human visual system appears to use. Second order priors penalize large second derivatives of depth (or disparity). When expressed as an energy minimization problem, the resulting energy has almost invariably been minimized by what Scharstein *et al*. [27] describe as "local methods": gradient descent [28] or PDE-based techniques such as level sets [11]. However, local methods struggle with (a) the long-range interactions associated with occlusion reasoning and (b) weak or multi-modal data likelihoods.

The introduction of "global" methods for energy minimization gave a considerable improvement in stereo reconstruction performance. Methods such as graph cuts [8, 20, 31] can find strong (although not global) optima of energies with long-range interactions. However, these methods have not previously incorporated second-order smoothness terms. This is despite the fact that the graph constructions necessary to include these terms are known [21]: the triple cliques which represent the second order terms are decomposed into several pairwise cliques and auxiliary nodes are added. Boykov and Veksler [7] say that "the allowed form of these triple cliques is very limited", but the *construction* is valid for any energy—the limitation is that the resulting graph is non-submodular, meaning that efficient methods of finding the global optimum were not known. In this paper we adapt a newly introduced optimizer [26], the "QPBO" method of Boros, Hammer and co-workers [5, 14], to compute good optima of the energy. Although these are not guaranteed to be global optima, our experiments show that by careful parametrization of the problem, good local optima can be found reliably.

Previous stereo algorithms have implemented approximations to second order smoothness priors. Ogale and Aloimonos [25] propose a "slanted scanline" algorithm, in which straight, 3d line segments are fitted to 2d image scanlines using an optimization method. This approach models visibility using an explicit uniqueness constraint, but the method is limited to image pairs, and the between-scanline optimization is local. Li and Zucker [23] introduce priors on slanted and curved surfaces, encouraging the second and third derivatives of depth to be zero. This novelly allows for curved surfaces in the solution, as shown in figure 1(c), and significantly improves on the fronto-parallel assumption on scenes where that assumption is violated. However their algorithm precomputes local surface normals and in fact optimizes a first-order prior on the normals, rather than a second-order prior on the disparities. Indeed, they discuss the global optimization of a second order prior, and conclude that this "makes the problem computationally infeasible".

A related class of methods is "segment-based" stereo. Early examples of this technique were proposed by Birchfield and Tomasi [2] and Tao *et al*. [29]. While their two approaches differ somewhat, both enforce the constraint that segmented regions of the image be planar, a trait common to the sequence of algorithms that succeeded Tao's [4, 15, 17, 35], which have shown excellent results on the "Middlebury " test set [27]. They all share the same three stage process—produce an over-segmentation of the reference image, generate a set of planar hypotheses for each segment, and optimize over the hypotheses—differing only in their implementation of each stage. Lin and Tomasi [24] explicitly minimize an energy including a second order prior, but are restricted to a local gradient-based optimization strategy where segmentation and depth estimation are interleaved. In many of these segment-based methods the assumption of the local planarity of scenes is not a general smoothness prior, but a hard constraint, which does not permit curved surfaces even when the data supports this. In contrast, we show that the method proposed here is an effective regularizer over both planar and curved surfaces.

2. Problem statement

Before describing this paper's main contribution, let us define the stereo problem, and the energy formulation that we propose to minimize.

The input is a set of $N+1$ images $\{\mathcal{I}_i\}_{i=0}^{N}$. The goal is to determine the dense disparity map, $\mathcal{D}$, of one *reference view*, say $\mathcal{I}_0$. A 2D vector, $\mathbf{x}$, denotes a pixel location in the reference view, the color of which is written as $I_0(\mathbf{x})$, and the corresponding disparity is $D(\mathbf{x})$. We are also given projection functions $\{\pi_i(\mathbf{x}, d) : \mathbb{R}^2 \mapsto \mathbb{R}^2\}_{i=1}^{N}$, where $\pi_i(\mathbf{x}, d)$ is the projection into image i of the 3D point corresponding to disparity (1/depth) d in front of pixel $\mathbf{x}$ in the reference view. For a rectified stereo pair, $N = 1$ and only π_1 is required, with the simple definition $\pi_1(\mathbf{x}, d) = \mathbf{x} + [d, 0]$.

The abbreviation $I_i^{\pi}(\mathbf{x}, d) = I_i(\pi_i(\mathbf{x}, d))$ will be used to reduce clutter, and may be read as "the color of the pixel corresponding to $\mathbf{x}$ in image i if the disparity at $\mathbf{x}$ is d".

The energy function to be minimized is a function of the disparity map $E(\mathcal{D})$, and is the sum of two terms: photo-consistency E_{photo}, which incorporates geometrical visibility reasoning, and smoothness E_{smooth}, as follows.

$$E(\mathcal{D}) = E_{\text{photo}}(\mathcal{D}) + E_{\text{smooth}}(\mathcal{D}). \tag{1}$$

The components of the energy shall now be described.

2.1. Data term

The data term in this paper is a standard photoconsistency term of the form

$$E_{\text{photo}}(\mathcal{D}) = \sum_{\mathbf{x}} \sum_{i=1}^{N} f\left(I_i^{\pi}(\mathbf{x}, D(\mathbf{x})) - I_0(\mathbf{x}), V_{\mathbf{x}}^{i}\right) \tag{2}$$

where $V^i_{\mathbf{x}}$ is a *visibility flag*, to be discussed below, indicating whether the 3D point defined by $(\mathbf{x}, D(\mathbf{x}))$ is visible in image i. Given $V^i_{\mathbf{x}}$, the consistency metric f is defined as

$$f(\Delta I, V) = \begin{cases} \rho_{\mathrm{d}}(\Delta I) & \text{if } V = 1 \\ \nu & \text{if } V = 0 \end{cases} \tag{3}$$

Here ν is the penalty cost paid by occluded pixels, and ρ_{d} is a robust measure of color difference, defined by

$$\rho_{\mathrm{d}}(I) = -\log\big(1 + \exp(-\|I\|^2/\sigma_{\mathrm{d}})\big), \tag{4}$$

where σ_{d} is set from the noise level in the sequence.

The **visibility** flag $V^i_{\mathbf{x}}$ adds nonlocal terms to the energy, making global optimization of this energy difficult, even before priors are incorporated. It is more correctly written $V_i(\mathbf{x}, \mathcal{D})$, indicating the dependence on many entries of the disparity map $\mathcal{D}$. We use the asymmetrical occlusion model of Wei and Quan [31], which reduces the complexity of the symmetrical multi-view occlusion model introduced in [20] from $\mathcal{O}(N)$ to $\mathcal{O}(1)$. This model adds pairwise terms to the energy, between nodes which are on the same epipolar lines. However the approximations made in [31] in order to ensure submodularity are unnecessary, given our optimization framework. As shall be seen in §3, optimization of the continuous $E(\mathcal{D})$ is expressed as a sequence of binary subproblems. It then becomes valuable to compute the decomposition into pairwise terms independently for each binary subproblem. This confers the advantage that the that the number of potentially occluding pixels is relatively small for each such subproblem, so the cost of including visibility is relatively low.

2.2. Surface smoothness

The smoothness prior places a cost, $\rho_{\mathrm{s}}(\cdot)$, on the smoothness $S(\cdot)$ of a neighborhood, $\mathcal{N}$, of pixels. In addition, a per-neighborhood conditional random field (CRF) weight $W(\mathcal{N})$, as discussed in §4, will be applied. E_{smooth} is the sum of smoothness costs over a defined set of pixel neighborhoods, $\mathbb{N}$, thus

$$E_{\mathrm{smooth}}(\mathcal{D}) = \sum_{\mathcal{N} \in \mathbb{N}} W(\mathcal{N}) \rho_{\mathrm{s}}(S(\mathcal{N}, \mathcal{D})). \tag{5}$$

with $\rho_{\mathrm{s}}(s) = \min(\sigma_{\mathrm{s}}, |s|)$, the truncated linear kernel.

Second-order priors are defined on three-pixel neighborhoods, and approximate the second derivative of disparity, thus:

$$S(\{\mathbf{p}, \mathbf{q}, \mathbf{r}\}, \mathcal{D}) = D(\mathbf{p}) - 2D(\mathbf{q}) + D(\mathbf{r}) \tag{6}$$

where the neighborhoods, $\mathcal{N} = \{\mathbf{p}, \mathbf{q}, \mathbf{r}\}$, are from the set of all 3×1 and 1×3 patches in the reference image. This function increases monotonically as the neighborhood diverges from collinearity, in contrast to the first-order prior traditionally used, $S(\{\mathbf{p}, \mathbf{q}\}, \mathcal{D}) = D(\mathbf{p}) - D(\mathbf{q})$, which increases monotonically as the neighborhood diverges from fronto-parallel.

3. Optimization

The above defines $E(\mathcal{D})$ as a function of a real-valued disparity image $\mathcal{D}$. In this section we describe how this energy is minimized, following recent generalizations of α-expansion [22, 26, 33]. In order to optimize the energy over the real-valued space, we reduce it to a sequence of binary problems as follows. Suppose we have a current estimate of the disparity, $\mathcal{D}_t$, and a *proposal* depth map $\mathcal{D}^{\mathrm{p}}$. In the α-expansion method, for example, the proposal depth at each step is a fronto-parallel plane [8]; in this paper we shall use more complex proposals (see §3.3). The goal is to optimally combine ("fuse") the proposal and current depth maps to generate a new depth map $\mathcal{D}_{t+1}$ for which the energy $E(\mathcal{D}_{t+1})$ is lower than $\mathcal{D}_t$. This *fusion move* is achieved by taking each pixel in $\mathcal{D}_{t+1}$ from one of $(\mathcal{D}_t, \mathcal{D}^{\mathrm{p}})$, as controlled by a binary indicator image $\mathcal{B}$ with elements $B(\mathbf{x})$:

$$\mathcal{D}^b(\mathcal{B}) = (1 - \mathcal{B}) \cdot \mathcal{D}_t + \mathcal{B} \cdot \mathcal{D}^{\mathrm{p}}, \tag{7}$$

where dot indicates elementwise multiplication. Thus, $\mathcal{B}$ may be read as "copy the disparity from the proposal $\mathcal{D}^{\mathrm{p}}(\mathbf{p})$ if $B(\mathbf{p}) = 1$, otherwise keep the current estimate $\mathcal{D}_t$". Then the energy $E(\mathcal{D})$ is a function only of the indicator image B, so we may define

$$\mathcal{D}_{t+1} = \mathcal{D}^b\left(\operatorname*{argmin}_{\mathcal{B}} E(\mathcal{D}^b(\mathcal{B}))\right) \tag{8}$$

This boolean optimization problem is then represented as a graph-cut problem, as described in §3.2 below. This will in general lead to a non-submodular graph, but we can use Quadratic Pseudo-Boolean Optimization (QPBO) [5, 14, 26], which is able to optimize non-submodular energies. Unlike the submodular case, where the global minimum $\mathcal{B}$ is guaranteed, QPBO returns a solution $\mathcal{B}$ and an associated mask $\mathcal{M}$ with the guarantee that at pixels $\mathbf{x}$ where $M(\mathbf{x}) = 1$, the value $b(\mathbf{x})$ is at the value it would have at the global minimum[1], but pixels where $M(\mathbf{x}) = 0$ have "unlabeled" values. By forcing $B(\mathbf{x}) = 0$ at those pixels, we ensure that $E(\mathcal{D}_{t+1}) \leq E(\mathcal{D}_t)$ (a result of the "autarky" property of QPBO [26, page 2]), thus guaranteeing not to increase the energy with each proposal.

Although in principle one could optimize our energy just using the above algorithm, in practice convergence would be slow, as the the number of unlabeled pixels at each fusion step may be high. In the next two sections we discuss three

[1] More correctly, *a* global optimum as there may be several labelings with the same energy

important procedures which can be used to greatly improve the performance of the algorithm: (1) a variety of alternative fusion moves; (2) the graph construction which allows each binary subproblem to be effectively solved; and (3) the selection of proposal depth maps.

3.1. Alternative fusion strategies

The fusion move described above states how to choose values for the binary labeling $\mathcal{B}$ at pixels unlabeled by QPBO. A number of alternatives are possible, as outlined below. In each case we use the labels 0 and 1 to represent the current and proposed solution, $\mathcal{D}_t$ and $\mathcal{D}^{\mathrm{p}}$, respectively. Many of these alternatives have appeared before in the literature, but we introduce two new strategies which give noticeable improvements in our results, and may prove useful in other contexts.

QPBO-F. *Fix to current* [33]: fix unlabeled nodes to 0, the current best labeling.

QPBO-L. *Lowest energy label* [22]: fix unlabeled nodes collectively to whichever of 0 or 1 gives the lowest energy.

QPBOP. *Probe*: probe the graph, as described in [6, 26], in order to find the labels of more nodes, that form part of an optimal solution.

QPBOI-F. *Fix to current and improve*: fix unlabeled nodes to 0, and transform this labeling using QPBOI, as described in [26].

QPBO-R. *Lowest cost label per region* (new approach, based on the "optimal splice" technique of [32]): split unlabeled nodes into *strongly connected regions* (SCRs), as per [1]. For each SCR, independently select the labeling, 0 or 1, which gives the lowest total energy for cliques connected to that region.

QPBOI-R. *Improve lowest cost label per region* (new approach): Label nodes as per QPBO-R, then use QPBOI to transform this labeling.

In §5.1 we empirically compare the various fusion strategies in the context of our problem.

3.2. Graph construction

As mentioned above, the conversion of the large-clique energy $E(\mathcal{D})$ into an equivalent pairwise representation is delayed until the binary optimization stage. Figure 2 demonstrates the construction of the graph used in each binary optimization, for a 1×3 pixel image. The graph contains only pairwise terms represented by the lines in the figure, linking the nodes. The black lines represent the data costs of equation (3), giving $2nN$ edges for an n pixel reference image. The blue lines are infinite edge costs which enforce the visibility constraint of the same equation, as per [31]; the line shown indicates that one (or both) of the disparity labels for pixel $\mathbf{p}$ occludes pixel $\mathbf{r}$ at disparity d_0 in $\mathcal{I}_1$. The list of pixel occlusion interactions is computed prior

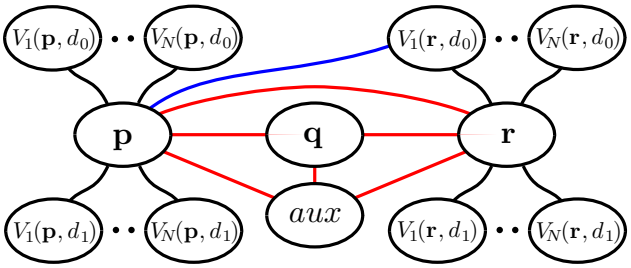


Figure 2. **Graph construction**. A graphical representation of the energy graph we construct for a 3×1 pixel image. Ovals represent nodes of the graph, and lines (edges) represent pairwise energy terms. Nodes $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ are binary variables encoding the disparities, (d_0, d_1), of those pixels. The nodes $V_1(\mathbf{p}, d_0)$, *etc*. encode whether (by way of example) pixel $\mathbf{p}$ is visible at disparity label 0 (i.e. disparity d_0) in $\mathcal{I}_1$; note that some of these nodes have been excluded for clarity. Black lines represent the data costs, blue lines the visibility constraint, and red lines the smoothness prior.

to solving the graph, and, while the list length is variable, it tends to be around nN edges.

The six red lines, which represent the smoothness costs of equation (5) for the only complete neighborhood, $\mathcal{N} = \{\mathbf{p}, \mathbf{q}, \mathbf{r}\}$, show how one triple clique is decomposed into six pairwise cliques, and an extra, latent node (labeled aux), using the decomposition described in [21]; note again that while the decomposition was originally given with regard to submodular graphs, it holds for any triple clique.

Graph complexity With the addition of a fourth pixel, $\mathbf{s}$, to create a 1×4 pixel image, the neighborhood $\{\mathbf{q}, \mathbf{r}, \mathbf{s}\}$ will share the edge $\mathbf{qr}$ with $\{\mathbf{p}, \mathbf{q}, \mathbf{r}\}$; therefore, generally, the total number of edges for a bidirectional, second-order smoothness prior, ignoring boundary effects, is $10n$, up from $2n$ for a first-order prior. It can therefore be seen that the use of a second, rather than first, order prior increases the graph size (number of edges) by a factor of approximately $(10 + 3N)/(2 + 3N)$—around 160% larger with two input images ($N = 1$), but only 60% larger with 5 input images. A similar analysis on the degree (number of incident edges) of each pixel node shows an increase of a factor of $(12 + 3N)/(4 + 3N)$, or about 114% higher for two images.

3.3. Proposal generation

The final component of the algorithm to be defined is the choice of proposals. In previous work [22, 33], the proposals have just been fronto-parallel planes (denoted "Same-Uni" below). As shown in [8], repeated fusion of these proposals leads to a strong local optimum in the submodular case. In the non-submodular case, the nature of these proposal disparity maps has a large effect on the generated disparity map, as we show empirically in §5. We use the following schemes for generating the j^{th} proposal disparity

map $\mathcal{D}_j^{\mathrm{p}}$:

SameUni Draw d_j from a uniform distribution, and set $D_j^{\mathrm{p}}(\mathbf{x}) = d_j$ for all $\mathbf{x}$.

SegPln Uses the ad-hoc approach of segmentation-based methods [17, 35] to generate a set of piecewise-planar proposals, which are then cycled through continuously. In this implementation, demonstrated in figure 3, the first stage of proposal generation involves a local window matching process [27] to generate an approximate (very noisy) disparity map. We then use two different image segmentation algorithms, one color-based [10], and one texture-based [12], and 14 sets of parameters in total, to generate segmentations of $\mathcal{I}_0$, ranging from highly under-segmented to highly over-segmented. For each segment in each segmentation we use LO-RANSAC [9] to find the plane that produces the greatest number of inlying correspondences from the first stage (given a suitable distance threshold), and set all the pixels in the segment to lie on that plane.

Smooth $D_j^{\mathrm{p}}(\mathbf{x}) = (D_j(\mathbf{x}+\Delta) + D_j(\mathbf{x}-\Delta))/2$, where $\Delta = [0,1]$ when j is odd, and $\Delta = [1,0]$ when j is even.

These proposal methods represent the different approaches used by the main types of stereo algorithms: the fronto-parallel proposals of SameUni are essentially those used at each iteration of an α-expansion-based stereo algorithm (except drawn from a continuous, rather than discrete, space); SegPln proposals are those used by segment-based algorithms; Smooth proposals, generated by a smoothing operation on the current disparity map, can be viewed as a proxy for local methods such as gradient descent. With QPBO-based fusion, we gain the benefits of all these algorithms—indeed, any stereo algorithm available—without affecting the global optimum. For example, the SegPln proposals, the main workhorse of our algorithm, are produced with a range of algorithms and parameter settings; in general we expect these disparity maps to be correct in some parts of the image, and for some parameter settings, but that no settings can be found for which any algorithm works best. By fusing the proposals in a well-defined energy minimization framework, the parameter sensitivity of these methods is turned into an advantage: we can select the best parts from each proposal, at the pixel (as opposed to segment) level.

4. Implementation

Some further implementation notes will allow the reader to more accurately replicate our method.

We normalize the range of disparities searched over for a particular image sequence to $[0,1]$ prior to the evaluation of E_{smooth}, in order to make our objective function invariant to image baseline, camera calibration and depth of field. The initial depth map, $\mathcal{D}_0$, is set to $D_0(\mathbf{x}) = \mathrm{rand}[0,1]$ for each $\mathbf{x}$ independently.

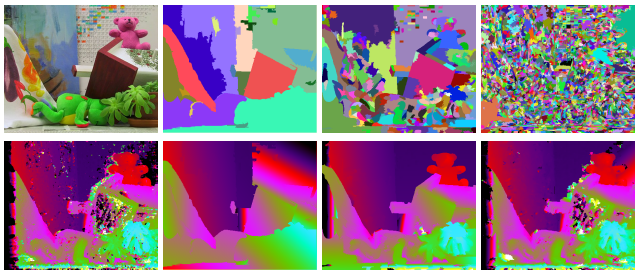


Figure 3. **SegPln proposal generation**. *Top row:* $\mathcal{I}_0$, and 3 of its 14 segmentations. *Bottom row:* approximate disparity map from window matching, and 3 SegPln proposals generated by fitting planes to each segment in the above segmentations.

Optimization is halted either when a maximum number of iterations, t_{max}, is reached, or when the average decrease in energy over the last 20 iterations drops below some threshold, $\delta E_{\mathrm{thresh}}$, whichever occurs first.

We use Kolmogorov's [19] implementations of QPBO, QPBOP and QPBOI. Both QPBOP and QPBOI methods make use of tree-recycling [18] for a fast implementation; the number of graph solves is at most linear in the number of unlabeled nodes for QPBOI, but exponential for QPBOP, though it should be noted that QPBOP labels nodes optimally, rather than approximately, as with QPBOI.

CRF weights The CRF weights $W(\cdot)$ are set to encourage disparity edges to align with edges in the reference image I_0. We generate a single mean-shift segmentation of the reference image ([10], $h_s = 4$ and $h_r = 5$), and assign one of two weights to each neighborhood, depending on whether or not it overlaps a segmentation boundary. Precisely, if L is the map which assigns to each pixel its segmentation label, then

$$W(\mathcal{N}) = \begin{cases} \lambda_{\mathrm{h}} & \text{if } L(\mathbf{p}) = L(\mathbf{q})\ \forall\ \mathbf{p},\mathbf{q} \in \mathcal{N} \\ \lambda_{\mathrm{l}} & \text{otherwise.} \end{cases} \tag{9}$$

Parameters We use the same parameter settings for all examples, i.e. $\nu = 0.01, \sigma_{\mathrm{d}} = 30C, \lambda_{\mathrm{l}} = 9N, \lambda_{\mathrm{h}} = 108N, \sigma_{\mathrm{s}} = 0.02$, where C is the number of color channels per input image. These settings were obtained by visual evaluation of a small number of Middlebury images (although it must be emphasised that they were not chosen with any reference to the Middlebury evaluation score). The order of the prior was found not to change the relative performance of parameter sets significantly.

5. Experiments

In this section we describe the experiments we carried out in evaluating the efficacy of QPBO in optimizing our

non-submodular energy, the trade-offs of each of the QPBO labeling methods, the effect of using different disparity proposals, and comparing our method, with its second-order prior, to the same method with a first-order prior, and other, competing approaches to stereo.

The optimization method used in each experiment is characterized by the order of the prior ("1op" for first-order prior, *etc.*), the set of proposals, the fusion strategy and the convergence criterion used to stop the optimization, *e.g.* "2op, SameUni, QPBOI-R, $\delta E_{\text{thresh}} = 0.01\%$", or "1op, SegPln, QPBOP, $t_{\max} = 200$".

5.1. Unlabeled nodes

The proportion of pixels that are labeled by QPBO has a direct impact on the quality of the solution found—trivially, if no nodes are labeled then (using QPBO-F) the final solution will be the same as the initial solution. We therefore ran experiments on the 4 Middlebury test sequences to evaluate what proportion of pixels were labeled, and which of the fusion strategies performed best at fixing these pixels. The results of these experiments are shown in figure 6. Figure 6(a) indicates that using SegPln proposals with a second-order prior generates the most unlabeled nodes for our chosen smoothness parameters, at 15% on average; we therefore used these settings to compare fusion strategies. Figure 6(b) shows that QPBOP rapidly becomes several orders of magnitude slower as the number of unlabeled pixels rises, while other methods roughly double in time over the same range; of these there is only fractional difference in speed, though order of fastest to slowest is consistently QPBO-F, QPBO-L, QPBO-R, QPBOI-F, QPBOI-R. In terms of energy reduction performance, QPBOP, which gives an optimal solution, performs best, while QPBO-F, with the simplest labeling strategy, performs worst. Figure 6(a) shows how the other strategies perform relative to these two, and indicates that QPBOI-R achieves the largest energy reduction. Considering the trade-off between time and efficacy, we found QPBOI-R to be the most suitable method for our problem, and used this in all further experiments.

5.2. Proposals

We applied all our proposal sets separately to each of our test sequences, with both first and second order priors, using QPBOI-R, $\delta E_{\text{thresh}} = 0.01\%$. However, as the Smooth proposal only performs well when applying it to an approximately correct disparity map, we prefixed the proposal set with the disparity maps generated using the other two proposal schemes, and repeated the set every six iterations, calling this set "Smooth*".

Figure 4 shows the results on the Middlebury "Venus" sequence. It can be seen that the fronto-parallel SameUni proposals generate generally piecewise-fronto-parallel solutions with both priors, while the piecewise-planar SegPln

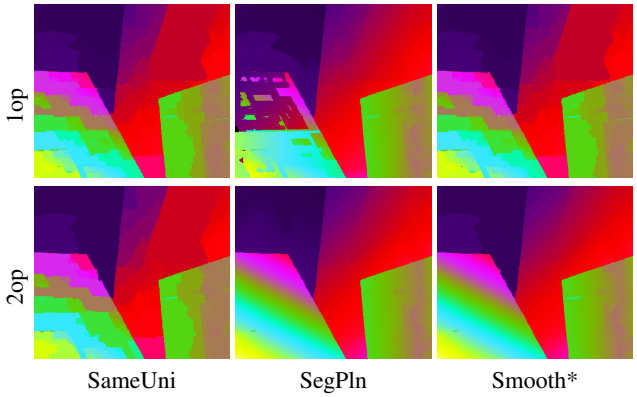


Figure 4. **Effect of proposals**. Output of our stereo method on the Venus sequence, using first and second order priors with our 3 proposal strategies.

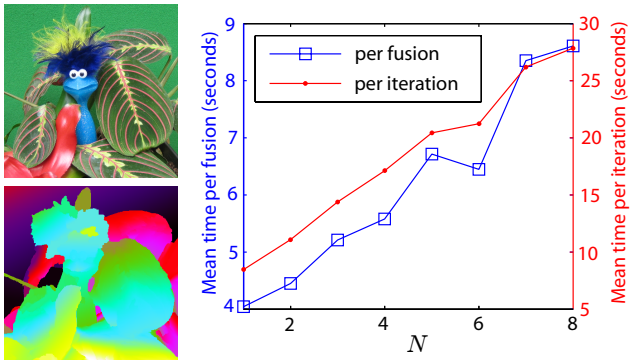
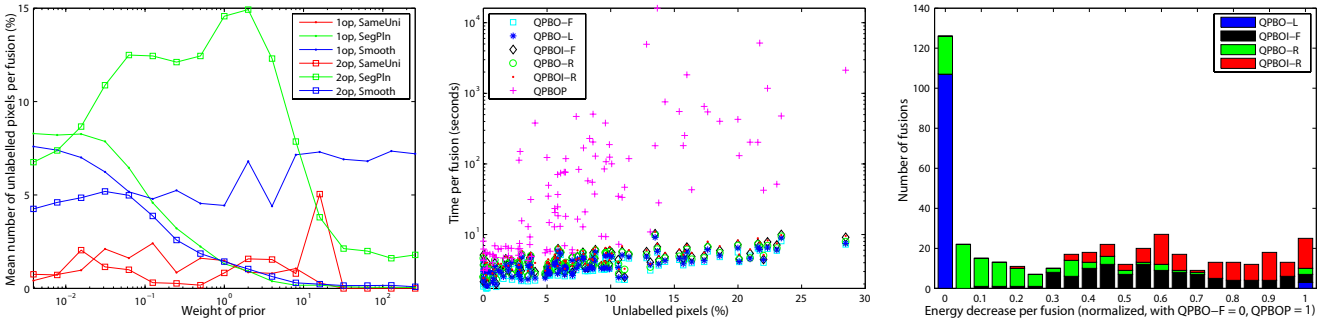


Figure 5. **Multiple, arbitrary views**. *Top left:* $\mathcal{I}_0$ for the "Plant & toy" sequence, which has arbitrary input views. *Bottom left:* Output disparity using "2op, Smooth*, QPBOI-R, $\delta E_{\text{thresh}} = 0.01\%$", for $N = 2$. *Right:* Graph of fusion (QPBO only) and iteration (including image graph construction) times as a function of N.

proposals generate piecewise-planar solutions, but with the first-order prior tending to favor more fronto-parallel surfaces over the correct solution. When these solutions are combined in the Smooth* proposal set, the first-order prior favors the SameUni solution, while the second-order prior favors the SegPln solution.

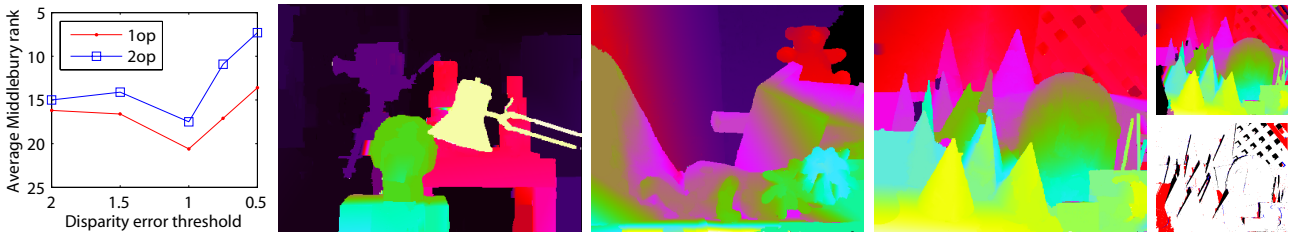
5.3. Second vs first order

We used the Middlebury stereo evaluation framework to compare the accuracy of results using first and second order priors. In order to remove biasing caused by proposal schemes (seen in the previous section) we only compare priors using Smooth* proposals (and QPBOI-R, $\delta E_{\text{thresh}} = 0.01\%$). Figure 7(a) shows the relative performance of the two priors, in terms of average rank in the Middlebury performance table. The graph shows that, not only does the second-order prior perform better at all error thresholds, but also that its performance improves more than the first order prior at the high-accuracy thresholds, relative to other algorithms, indicating improved subpixel accuracy. This ef-



(a) (b) (c)

Figure 6. **Proposal and labeling methods**. (a) Graph showing proportion of unlabeled pixels per fusion with different priors, prior weightings (*i.e.* multiplicative factor on λ_{l} and λ_{h}) and proposal methods. (b) Graph showing the time per fusion as a function of unlabeled pixels and fusion strategy. (c) Stacked histogram showing the effect of fusion strategies on energy, relative to QBPO-F and QPBOP.



(a) (b) (c) (d) (e)

Figure 7. **Middlebury performance**. (a) A graph of average rank on the Middlebury stereo evaluation, against disparity error threshold, for both first and second order priors. (b)–(d) Output disparity using "2op, Smooth*, QPBOI-R, $\delta E_{\mathrm{thresh}} = 0.01\%$", for the Middlebury "Tsukuba", "Teddy" and "Cones" sequences respectively. (e) *Top:* Output disparity using the same method as (d), but with no visibility constraint (*i.e.* $V_i(\mathbf{x}) = 1\ \forall \mathbf{x}$). *Bottom:* Visibility map for $\mathcal{I}_1$ of "Cones"—pixels deemed occluded according to the following disparity maps are painted (covering the previous color) in the following order: disparity map above, red; (d), blue; ground truth, black.

fect can be explained by the fact that non-fronto-parallel planes, as well as curved surfaces, are better modeled by the second-order prior, as demonstrated in figure 8.

Figure 7(e) highlights the benefits of a visibility constraint (comparing numbers of red and blue pixels)—by reducing the number of falsely occluded pixels, it essentially encourages uniqueness of correspondences between input images. As unique correspondence is a constraint on real-world scenes, incorporating such a constraint in a stereo framework produces better results.

5.4. Multiple & arbitrary views

The formulation of our objective function allows for any number of input images to be used, and for those images to have arbitrary viewpoints. Figure 5 shows results for such a dataset—the "Plant & toy" sequence from [33]. We found little or no qualitative improvement between $N = 2$ (three views) and $N > 2$, something we believe can be attributed to the fact that three views are sufficient (in this case) to ensure that each pixel of $\mathcal{I}_0$ is visible in at least one other view. However, should more views be required, figure 5(*right*), shows that, in practice, the time per fusion iteration (with and without graph construction overheads such as image sampling and visibility computation) rises linearly with N.

6. Conclusion

This paper has shown that second-order smoothness priors can be incorporated into graph-cut based stereo reconstruction. This was not previously possible, because the non-submodular energies led to infeasibly complex optimizations. Previous stereo algorithms using second-order priors were limited by local optimizers. In particular, the combination of second-order priors with simultaneous global visibility reasoning was not possible. The paper's main contribution is a framework for optimizing the resulting objective function. We have demonstrated that this method produces depth maps that accurately reconstruct the scene at a subpixel level. The algorithm can be equally applied to multi-view stereo with arbitrary camera viewpoints, and does so at a computational cost linear in N.

An interesting feature of the optimization strategy, and in particular the "SegPln" proposals, is that it can make use of existing algorithms, which may sometimes be rather ad-hoc, and combine their results in a principled way. We expect this property to offer considerable opportunities for improvement of the basic method in terms of the quality of optima discovered, and the speed at which they can be found.

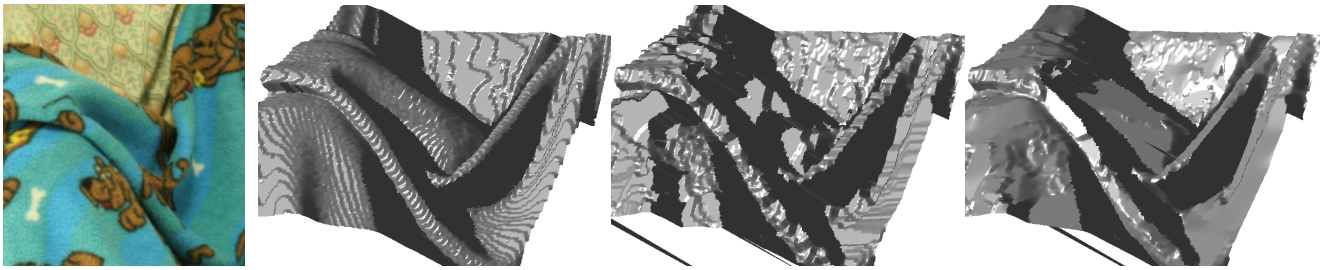


Figure 8. **Curved surfaces**. *Left to right:* $\mathcal{I}_0$ for the Middlebury "Cloth3" sequence, ground truth (discretized) disparity surface (3-d view of disparity map), disparity surfaces generated using Smooth* proposals and 1op and 2op respectively. (Spurious pixels have been fixed to the back-plane, for improved visualization.)

Acknowledgements We thank Vladimir Kolmogorov for making available his QPBO software, and also for discussing graph cut stereo with us. Research funded by EPSRC grants EP/C007220/1, EP/C006631/1(P) and a CASE studentship with Sharp.

References

[1] A. Billionnet and B. Jaumard. A decomposition method for minimizing quadratic pseudoboolean functions. *Operations Research Letters*, 8(3):161–163, Jun 1989.

[2] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *Proc. ICCV*, pages 489–495, Sep 1999.

[3] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, USA, Aug 1987.

[4] M. Bleyer and M. Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. In *Intl. Conf. Image Proc.*, volume 5, pages 2997–3000, Oct 2004.

[5] E. Boros, P. L. Hammer, and X. Sun. Network flows and minimization of quadratic pseudo-boolean functions. *Technical Report RRR 17-1991, RUTCOR Research Report*, May 1991.

[6] E. Boros, P. L. Hammer, and G. Tavares. Preprocessing of unconstrained quadratic binary optimization. *Technical Report RRR 10-2006, RUTCOR Research Report*, Apr 2006.

[7] Y. Boykov and O. Veksler. Graph cuts in vision and graphics: Theories and applications. In *The Handbook of Mathematical Models in Computer Vision*. Springer, 2006.

[8] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE PAMI*, 23(11):1222–1239, 2001.

[9] O. Chum, J. Matas, and Š. Obdržálek. Enhancing RANSAC by generalized model optimization. In *Proc. Asian Conf. on Computer Vision*, volume 2, pages 812–817, Jan 2004.

[10] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE PAMI*, 24(5):603–619, 2002.

[11] O. D. Faugeras and R. Keriven. Complete dense stereovision using level set methods. In *Proc. ECCV*, pages 379–393, 1998.

[12] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, Sep 2004.

[13] W. E. L. Grimson. *From Images to Surfaces: A Computational Study of the Human Early Visual System*. MIT Press, 1981.

[14] P. L. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 0-1 optimization. *Mathematical Programming*, 28:121–155, 1984.

[15] L. Hong and G. Chen. Segment-based stereo matching using graph cuts. In *Proc. CVPR*, pages 74–81, 2004.

[16] H. Ishikawa and D. Geiger. Rethinking the prior model for stereo. In *Proc. ECCV*, pages 526–537, 2006.

[17] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Proc. ICPR*, pages 15–18, 2006.

[18] P. Kohli and P. H. S. Torr. Efficiently solving dynamic Markov Random Fields using graph cuts. In *Proc. ICCV*, pages 922–929, 2005.

[19] V. Kolmogorov. Discrete MRF optimization software. `http://www.adastral.ucl.ac.uk/~vladkolm/-software.html`.

[20] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. ECCV*, volume 3, page 82, 2002.

[21] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE PAMI*, 26(2):147–159, 2004.

[22] V. Lempitsky, C. Rother, and A. Blake. LogCut - efficient graph cut optimization for Markov Random Fields. In *Proc. ICCV*, 2007.

[23] G. Li and S. W. Zucker. Surface geometric constraints for stereo in belief propagation. In *Proc. CVPR*, pages 2355–2362, 2006.

[24] M. Lin and C. Tomasi. Surfaces with occlusions from layered stereo. *IEEE PAMI*, 28(8):710–717, 2004.

[25] A. S. Ogale and Y. Aloimonos. Shape and the stereo correspondence problem. *IJCV*, 65(3):147–162, 2005.

[26] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *Proc. CVPR*, 2007.

[27] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.

[28] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. In *Proc. CVPR*, volume 1, pages 552–559, Jun 2004.

[29] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. In *Proc. ICCV*, pages 532–539, 2001.

[30] D. Terzopoulos. Multilevel computational processes for visual surface reconstruction. *Computer Vision, Graphics and Image Processing*, 24(1):52–96, Oct 1983.

[31] Y. Wei and L. Quan. Asymmetrical occlusion handling using graph cut for multi-view stereo. In *Proc. CVPR*, volume 2, pages 902–909, 2005.

[32] O. Woodford, I. D. Reid, P. H. S. Torr, and A. W. Fitzgibbon. Fields of experts for image-based rendering. In *Proc. BMVC.*, volume 3, pages 1109–1108, 2006.

[33] O. Woodford, I. D. Reid, P. H. S. Torr, and A. W. Fitzgibbon. On new view synthesis using multiview stereo. In *Proc. BMVC.*, volume 2, pages 1120–1129, 2007.

[34] O. J. Woodford. OJW's Image Based Rendering (IBR) Toolbox v2. `http://www.robots.ox.ac.uk/~ojw/software.htm`.

[35] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér. Stereo matching with color-weighted correlation, hierachical belief propagation and occlusion handling. In *Proc. CVPR*, pages 2347–2354, 2006.