Semi-Global Matching: a principled derivation in terms of Message Passing

Amnon Drory[1], Carsten Haubold[2], Shai Avidan[1], Fred A. Hamprecht[2]

[1]Tel Aviv University, [2]University of Heidelberg

Semi-global matching, originally introduced in the context of dense stereo, is a very successful heuristic to minimize the energy of a pairwise multi-label Markov Random Field defined on a grid. We offer the first principled explanation of this empirically successful algorithm, and clarify its exact relation to belief propagation and tree-reweighted message passing. One outcome of this new connection is an uncertainty measure for the MAP label of a variable in a Markov Random Field.

1 Introduction

Markov Random Fields (MRFs) have become a staple of computer vision. Practitioners appreciate the ability to provoke nonlocal effects by specifying local interactions only; and theoreticians like how easy it is to specify valid nontrivial distributions over high-dimensional entities. Unfortunately, *exact* maximum a posteriori (MAP) inference is tractable only for special cases. Important examples include binary[1] MRFs with only submodular potentials (by minimum st-cut [10]), multilabel MRFs with potentials that are convex over a linearly ordered label set (minimum st-cut [7]), tree-shaped MRFs (by dynamic programming), Gaussian MRFs (by solving a linear system) and non-submodular binary pairwise MRFs without unary terms defined over planar graphs (by perfect matching [15]). Many real-world problems have been cast as (and sometimes strong-armed to match) one of these special cases. Inference in more general MRFs, such as non-tree shaped multilabel MRFs, is NP-hard and heuristics such as alpha-expansion [1], damped loopy belief propagation [4] or tree-reweighted message passing [17, 9] are often used.

One heuristic that has become very influential, especially in the context of dense disparity from stereo, is Semi-Global Matching (SGM). It is trivial to implement, extremely fast, and has ranked highly in the Middlebury [16], and the KITTI benchmarks [3] for several years. While intuitive, this successful heuristic has not found a theoretical characterization to date.

The present work establishes, for the first time, the precise relation of SGM to non-loopy belief propagation on a subgraph of the full MRF, and to tree-reweighted message passing with parallel tree-based updates [17]. This allows SGM to be viewed in the light of the rich literature on these techniques, . Based on these insights, we propose a lower bound based uncertainty measure that may benefit downstream processing.

[1] in the sense that nodes can take one of two possible labels

2 Notation and Background

2.1 Markov Random Fields

Computer Vision problems are often cast in terms of MRFs. Throughout this paper we focus on *pairwise* MRFs, defined on an undirected graph $G=(V,\mathcal{E})$. Each node $\mathbf{p} \in V$ is a random variable which can take values $d_{\mathbf{p}}$ from a finite set of labels Σ. Here we assume that all variables have the same label space. Then a *labeling* $\mathbf{D} \in \Sigma^{|V|}$ is an assignment of one label for each variable. The best labeling $\mathbf{D}^*$ corresponds to the solution with the lowest energy $\mathrm{argmin}_{\mathbf{D}\in\Sigma^{|V|}} E(\mathbf{D})$, with

$$E(\mathbf{D}) = \sum_{\mathbf{p}\in V} \varphi_{\mathbf{p}}(d_{\mathbf{p}}) + \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{E}} \varphi_{\mathbf{p},\mathbf{q}}(d_{\mathbf{p}}, d_{\mathbf{q}}) \tag{1}$$

where $\varphi_{\mathbf{p}}(\cdot)$ is a *unary term* defined over the node $\mathbf{p}$, and $\varphi_{\mathbf{p},\mathbf{q}}(\cdot,\cdot)$ is a pairwise term defined over the edge $(\mathbf{p},\mathbf{q})$.[2]

2.2 Min-Sum Belief Propagation (BP)

Min-sum Belief Propagation [11, 18] is an efficient dynamic programming algorithm for exact energy minimization on MRFs whose underlying graph is a *tree*. It calculates each node's *energy min marginal (EMM)*, or *belief* $\beta_{\mathbf{p}}(d_{\mathbf{p}})$, by sending messages along the edges. As soon as node $\mathbf{p}$ has received messages from all its neighbors but node $\mathbf{q}$, it can send the following message to $\mathbf{q}$:

$$m_{\mathbf{p}\to\mathbf{q}}(d_{\mathbf{q}}) = \min_{d_{\mathbf{p}}\in\Sigma} \Big(\varphi(d_{\mathbf{p}}) + \varphi(d_{\mathbf{p}}, d_{\mathbf{q}}) + \sum_{(\mathbf{k},\mathbf{p})\in\mathcal{E},\mathbf{k}\neq\mathbf{q}} m_{\mathbf{k}\to\mathbf{p}}(d_{\mathbf{p}})\Big) \tag{2}$$

The belief for each node is calculated from all the messages entering into it:

$$\beta_{\mathbf{p}}(d_{\mathbf{p}}) = \varphi(d_{\mathbf{p}}) + \sum_{(\mathbf{k},\mathbf{p})\in\mathcal{E}} m_{\mathbf{k}\to\mathbf{p}}(d_{\mathbf{p}}) \tag{3}$$

If we consider a certain node $\mathbf{r}$ to be the root of the tree, then the BP algorithm can be divided into 2 parts. In the *inward pass* only the messages that are flowing inwards from the leaves towards $\mathbf{r}$ are calculated. Then, the rest are calculated in the *outward pass*. To calculate $\mathbf{r}$'s own belief, $\beta_{\mathbf{r}}(d_{\mathbf{r}})$, only the inward pass is necessary.

The belief $\beta_{\mathbf{p}}(\cdot)$ is also known as node $\mathbf{p}$'s EMM which means $\beta_{\mathbf{p}}(\ell)$ is the energy of the best labeling where node $\mathbf{p}$ takes label ℓ.

[2] We will drop the indices in $\varphi(d_{\mathbf{p}})$ and $\varphi(d_{\mathbf{p}}, d_{\mathbf{q}})$ as they are clear from the function inputs.

2.3 Tree-Reweighted Message-Passing (TRW-T)

Tree-reweighted message-passing with tree-based updates (TRW-T [17]) is an approximate energy minimization algorithm for general MRFs. It works by iteratively decomposing the original MRF into a convex sum of tree MRFs, optimizing each tree separately, and recombining the results. The result of each iteration is a *reparametrization* of the original MRF, which is a different MRF that has the same energy function E, but different unary and pairwise terms, φ.

TRW-T can be applied to an MRF with an underlying graph $G = (V, \mathcal{E})$, potentials φ, and a set of trees $\{T^1, \ldots, T^{N_\mathcal{T}}\}$ which are all sub-graphs of G, such that each node and edge in G belongs to at least one tree[3]. Each tree $T^i = (V^i, \mathcal{E}^i)$ has a weight ρ^i so that $\sum_{i=1}^{N_\mathcal{T}} \rho^i{=}1$. The *node weight* $\rho_\mathbf{p}$ is defined as the sum of weights of all trees that contain node $\mathbf{p}$. *Edge weights* $\rho_\mathbf{pq}$ are defined similarly.

In each iteration of TRW-T, the following steps are performed:

1. An MRF is defined for each tree by defining tree unary and pairwise terms for each node and edge, by:
$$\varphi^i(d_\mathbf{p}) = \frac{1}{\rho_\mathbf{p}} \varphi^G(d_\mathbf{p}), \qquad \varphi^i(d_\mathbf{p}, d_\mathbf{q}) = \frac{1}{\rho_\mathbf{pq}} \varphi^G(d_\mathbf{p}, d_\mathbf{q}) \tag{4}$$
where by φ^i we denote the unary and pairwise terms for the tree T^i, and by φ^G we denote the terms of the full MRF.
2. The *min-sum belief propagation* algorithm is performed on each tree T^i separately, generating an EMM $\beta^i(\mathbf{p})$ for each node $\mathbf{p}$. A reparametrization of the tree MRF is defined, where the unary terms are: $\widetilde{\varphi^i_\mathbf{p}} \triangleq \beta^i(\mathbf{p})$.
3. The reparametrizations of the tree MRFs are aggregated to produce a reparametrization for the graph-MRF. For the unary terms, this is done by:
$$\widetilde{\varphi^G_\mathbf{p}} = \sum_{i:\mathbf{p} \in V^i} \rho^i \widetilde{\varphi^i_\mathbf{p}} \tag{5}$$

It is hoped, though not guaranteed, that repeated iterations of this process will flow potential from the entire graph into the global unary terms $\widetilde{\varphi^G_\mathbf{p}}$, and turning them into *EMMs*. The final label map can then be found by choosing, for each node, the label that minimizes its unary term:

$$d^*_\mathbf{p} = \operatorname*{argmin}_{d_\mathbf{p} \in \Sigma} \widetilde{\varphi^G_\mathbf{p}}(d_\mathbf{p}) \tag{6}$$

It is important to note that there are no convergence assurances for TRW-T. This means that the energy of the labeling that is produced after each iteration is not guaranteed to be lower, or even equal, to the energy of the previous labeling. Kolmogorov's TRW-S [9] improves upon TRW-T and provides certain convergence guarantees, but can not be related to SGM directly.

[3] In the presentation of the algorithm in [17] *spanning* trees are used, but as mentioned in [9], this is not necessary

Tree Agreement and Labeling Quality Estimation From the description of the TRW-T process it follows that the graph-MRF's energy function E^G is a convex combination of the tree energy functions E^i, i.e. $E^G(\mathbf{D}) = \sum_{i=1}^{N_\mathcal{T}} \rho^i E^i(\mathbf{D})$. Recall we're interested in MAP inference, i.e. finding the labeling with the minimal energy:

$$\min_{\mathbf{D}\in\Sigma^{|V|}} E^G(\mathbf{D}) = \min_{\mathbf{D}\in\Sigma^{|V|}} \sum_{i=1}^{N_\mathcal{T}} \rho^i E^i(\mathbf{D}) \geq \underbrace{\sum_{i=1}^{N_\mathcal{T}} \min_{\mathbf{D}^i\in\Sigma^{|V|}} \rho^i E^i(\mathbf{D}^i)}_{E_{LB}} \tag{7}$$

We get a *lower bound* on the minimum graph energy by allowing each tree energy to be minimized separately. TRW-T iteratively computes this lower bound, denoted by E_{LB}, by running BP on each tree. Notice that the inequality becomes equality *iff* the best labelings of all trees agree on the label assigned to every node. If this occurs, TRW-T can stop and the labeling that is found is indeed the best labeling $\mathbf{D}^*$. The lower bound can be used for estimating how good a given labeling $\mathbf{D}$ is. For any labeling $\mathbf{D}$ it holds that $E(\mathbf{D}) \geq E(\mathbf{D}^*)$. From this it follows that $E(\mathbf{D}) \geq E_{LB}$. If the relative difference between $E(\mathbf{D})$ and E_{LB} is small, that could indicate that $\mathbf{D}$ is a good labeling. TRW's reparametrizations aim to adjust the unaries such that E_{LB} approaches $E(\mathbf{D}^*)$

2.4 Semi-Global Matching (SGM)

As part of the Semi-Global Matching algorithm [5] for dense stereo matching, Hirschmüller presented an efficient algorithm for approximate energy minimization for a grid-shaped pairwise MRF. It divides the grid-shaped problem into multiple one-dimensional problems defined on *scanlines*, which are straight lines that run through the image in multiple directions. Inference in each scanline is performed separately, and the results are then combined to produce a labeling (in stereo: disparity map) for the image.

We denote by $\mathbf{p}=(x, y)$ the location of a pixel, and by $\mathbf{r}=(dx, dy)$ the direction of a scanline. SGM's processing of a scanline consists of the following recursive calculation:

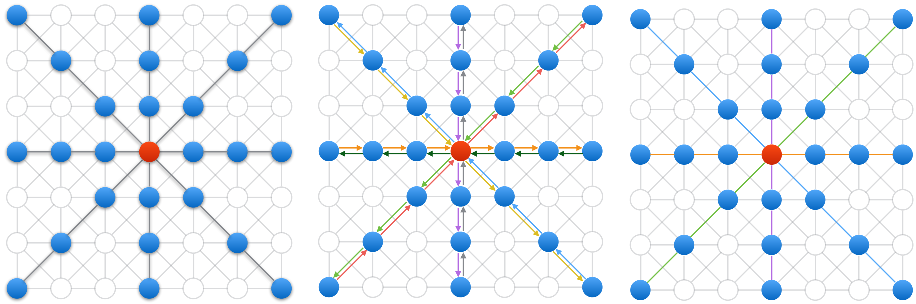
$$L^\mathbf{r}(d_\mathbf{p}) = \varphi(d_\mathbf{p}) + \min_{d_{\mathbf{p}-\mathbf{r}}\in\Sigma} \left\{ L^\mathbf{r}(d_{\mathbf{p}-\mathbf{r}}) + \varphi(d_{\mathbf{p}-\mathbf{r}}, d_\mathbf{p}) \right\} \tag{8}$$

We denote the set of all scanline directions by $\mathcal{R}$, and its size by $|\mathcal{R}|$, which is typically 8 or 16. The aggregation of scanline results is accomplished by simple summation:

$$S(d_\mathbf{p}) = \sum_{\mathbf{r}\in\mathcal{R}} L^\mathbf{r}(d_\mathbf{p}) \tag{9}$$

Finally, a pixel labeling is obtained by defining:

$$d_\mathbf{p}^{\mathrm{SGM}} = \operatorname*{argmin}_{d_\mathbf{p}\in\Sigma} S(d_\mathbf{p}) \tag{10}$$



(a) The *Union Jack*, a tree-shaped graph

(b) All the scanlines that go through the red pixel $\mathbf{p}$ (in 8 directions)

(c) The corresponding *Undirected Lines*

Fig. 1. The *Union Jack*, its 8 scanlines and 4 *Undirected Lines*. When performing inference for the red pixel, SGM ignores all the greyed-out nodes and edges of the original MRF.

3 Synthesis

The qualitative nature of the connection between SGM and BP as well as TRW-T may already have emerged in the foregoing. We now make it quantitative.

3.1 Relation between SGM and BP on Union Jack

Claim 1: $S(d_\mathbf{p})$, the effective label costs computed at a pixel $\mathbf{p}$ by 8-directional SGM are identical to the min-marginals computed on a *Union Jack* (see below) centered at pixel $\mathbf{p}$, up to (easily correctable) over-counting of the unary term at $\mathbf{p}$.

SGM's treatment of each scanline is similar to a forward pass of *min-sum belief propagation* (BP) on it, flowing messages from the first pixel on the scanline to the last. To be exact, SGM's processing of a scanline amounts to the *Forward Algorithm* [13], a close relative of min-sum BP. [4] The relation between the two algoriths is given by:

$$m_{(\mathbf{p}-\mathbf{r})\to\mathbf{p}}(d_\mathbf{p}) = L^\mathbf{r}(d_\mathbf{p}) - \varphi(d_\mathbf{p}) \tag{11}$$

In this section we focus on the common case where SGM is used with scanlines in 8 directions. For each pixel $\mathbf{p}$ we define the *Union Jack of* $\mathbf{p}$ as the union of all scanlines that pass through $\mathbf{p}$. The *Union Jack* is a tree with 8 branches that

[4] In terms of message passing on a *factor graph* , one variant represents factor-to-node messages, while the other gives node-to-factor messages

meet only at $\mathbf{p}$ itself, see Figure 1(a). By performing the *inward pass* of BP on the *Union Jack*, we can calculate $\beta(d_{\mathbf{p}})$, the EMM for $\mathbf{p}$ *on its Union Jack*:

$$\beta(d_{\mathbf{p}}) = \varphi(d_{\mathbf{p}}) + \sum_{\mathbf{r}\in\mathcal{R}} m_{(\mathbf{p}-\mathbf{r})\to\mathbf{p}}(d_{\mathbf{p}}) \tag{12}$$

We'll next show that the function $S(d_{\mathbf{p}})$ calculated by SGM for pixel $\mathbf{p}$ is almost identical to $\beta(d_{\mathbf{p}})$. We note that each message $m_{(\mathbf{p}-\mathbf{r})\to\mathbf{p}}(d_{\mathbf{p}})$ is calculated from a single scanline, on which equation (11) applies:

$$\begin{aligned}\beta(d_{\mathbf{p}}) &= \varphi(d_{\mathbf{p}}) + \sum_{\mathbf{r}\in\mathcal{R}} \Big(L^{\mathbf{r}}(d_{\mathbf{p}}) - \varphi(d_{\mathbf{p}})\Big) \\ &= \Big(\sum_{\mathbf{r}\in\mathcal{R}} L^{\mathbf{r}}(d_{\mathbf{p}})\Big) - (|\mathcal{R}|-1)\varphi(d_{\mathbf{p}}) = S(d_{\mathbf{p}}) - 7\varphi(d_{\mathbf{p}})\end{aligned} \tag{13}$$

From the last line we can see that the function $S(d_{\mathbf{p}})$ that SGM calculates for a pixel $\mathbf{p}$ is similar to $\mathbf{p}$'s EMM calculated on its *Union Jack*. The difference is that SGM *over-counts* the unary term $\varphi(d_{\mathbf{p}})$ 7 times. This can easily be fixed by adding a final subtraction stage to SGM.

In summary, it turns out that (over-counting corrected) SGM assigns to each pixel the label it would take in the best labeling *of its Union Jack*. This can be seen as an approximation of the label it would take in the optimal labeling *of the entire graph*. Unfortunately, finding that label would require calculating the EMM of each pixel relative to *the entire graph*, which generally is intractable.

Performing BP on each *Union Jack* separately would be inefficient, as many calculations are repeated in different *Union Jacks*. SGM avoids this by reusing the messages L_r computed on each scanline for all *Union Jacks* that contain it.

3.2 Relation between SGM and TRW-T

Claim 2: Performing over-counting corrected SGM on an MRF with energy

$$E = \sum_{\mathbf{p}\in V} \varphi(d_{\mathbf{p}}) + \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{E}} \varphi(d_{\mathbf{p}}, d_{\mathbf{q}}) \tag{14}$$

is equivalent to performing one iteration of TRW-T on an MRF with energy

$$\widetilde{E} = \sum_{\mathbf{p}\in V} \varphi(d_{\mathbf{p}}) + C_0 \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{E}} \varphi(d_{\mathbf{p}}, d_{\mathbf{q}}) \tag{15}$$

where $C_0 > 0$ is a constant, e.g. for 8-directional SGM, $C_0 = \frac{1}{4}$

The SGM algorithm closely resembles a single iteration of TRW-T. In both, we perform BP on trees, and then aggregate the results into an approximate belief for each pixel. Next we choose for each pixel the label that minimizes its own approximate belief.

Where SGM differs significantly from TRW-T is in the way the graph MRF's energy terms are used to define the tree MRF terms. SGM simply copies the terms, while TRW-T weighs them as to make the graph terms a convex sum of the tree terms (equations (4)). This reveals that SGM is not directly equivalent to a single iteration of TRW-T on the same MRF. Instead, performing SGM on a specific MRF is equivalent to performing one iteration of TRW-T on an MRF whose pairwise terms are multiplied by a constant C_0 (equation (15)).

At first glance, the conclusion from claim 2 seems to be that SGM is in fact different from TRW-T, since applying both algorithms to the same MRF will generally produce different results. However, in a real-life setting this is not necessarily the case. In many application scenarios, the energy function is defined as $E = E_{Data} + \lambda E_{Smoothness}$, and λ is learned from a training set. In these cases, the difference between the two algorithms will disappear, since a different value of λ will be learned for each, in such a way that $\lambda_{TRW} = C_0 \lambda_{SGM}$.

To fully define TRW-T we must specify the set of trees $\mathcal{T}$ that we use for TRW-T, and their weights ρ. We define $\mathcal{T}$ to be the set of SGM's *Undirected Lines*, with equal weights ρ_0 for all trees. An *Undirected Line* (UL) is simply a scanline, except no direction of traversal is defined. Each UL corresponds to two opposite-directed scanlines, $\mathbf{r}$ and $\bar{\mathbf{r}}$ (see figure 1(b)-(c)). Using equation (11), we can show that performing BP on the UL is very closely related to running SGM on both $\mathbf{r}$ and $\bar{\mathbf{r}}$:

$$\beta^{\mathbf{r}}(d_{\mathbf{p}}) + \varphi(d_{\mathbf{p}}) = L^{\mathbf{r}}(d_{\mathbf{p}}) + L^{\bar{\mathbf{r}}}(d_{\mathbf{p}}) \tag{16}$$

where $\beta^{\mathbf{r}}(d_{\mathbf{p}})$ is the belief calculated for pixel $\mathbf{p}$ by running BP on the UL corresponding to scanline $\mathbf{r}$. SGM's output $S(d_{\mathbf{p}})$ can be explained as a sum of the beliefs calculated for pixel $\mathbf{p}$ on each UL passing through it, with some over-counting of the unary term:

$$S(d_{\mathbf{p}}) = \sum_{\mathbf{r}\in\mathcal{R}} L^{\mathbf{r}}(d_{\mathbf{p}}) = \underbrace{\sum_{\mathbf{r}\in\mathcal{R}^+} \beta^{\mathbf{r}}(d_{\mathbf{p}})}_{\widetilde{S}(d_{\mathbf{p}})} + \frac{|\mathcal{R}|}{2}\varphi(d_{\mathbf{p}}) \tag{17}$$

where $\mathcal{R}^+$ is the group of ULs, and $\widetilde{S}(d_{\mathbf{p}})$ is the output of an over-counting corrected version of SGM.

Proof of Claim 2: We wish to show that the labeling created by SGM for an MRF with energy E (equation (14)) is the same as the labeling created by one iteration of TRW-T on an MRF with energy $\widetilde{E}$ (equation (15)). Let us first introduce a scale factor $C_1 > 0$ to define $\widehat{E} = C_1\widetilde{E}$. $\widehat{E}$ and $\widetilde{E}$ have the same labeling after an iteration of TRW-T, as the argmin is unaffected by scaling (equation (6)). We will now show how to choose C_0 and C_1 to make the tree MRFs equal. Note that SGM simply copies the potentials $\varphi = \varphi^G$, whereas TRW-T weighs them according to equation (4). Here $\rho_{\mathbf{p}} = \rho_0\frac{|\mathcal{R}|}{2}$, and $\rho_{\mathbf{pq}} = \rho_0\epsilon$, where ϵ is the number of ULs in which each edge appears (for 8 directional SGM, $\epsilon = 1$).

The unary potential $\hat{\varphi}^i(d_\mathbf{p})$ in a tree of TRW-T applied to $\widehat{E}$ is then given as

$$\hat{\varphi}^i(d_\mathbf{p}) = \frac{1}{\rho_\mathbf{p}}\hat{\varphi}^G(d_\mathbf{p}) = \frac{1}{\rho_\mathbf{p}}C_1\varphi^G(d_\mathbf{p}). \tag{18}$$

Thus, to make $\hat{\varphi}^i(d_\mathbf{p}) = \varphi(d_\mathbf{p})$ hold, we need to choose $C_1 = \rho_\mathbf{p} = \rho_0\frac{|\mathcal{R}|}{2}$. This leaves C_0 to be determined by the pairwise potential:

$$\hat{\varphi}^i(d_\mathbf{p}, d_\mathbf{q}) = \frac{1}{\rho_\mathbf{pq}}C_0\hat{\varphi}^G(d_\mathbf{p}, d_\mathbf{q}) = \frac{1}{\rho_\mathbf{pq}}C_1C_0\varphi^G(d_\mathbf{p}, d_\mathbf{q}) \tag{19}$$

$$= \frac{\rho_0\frac{|\mathcal{R}|}{2}}{\rho_0\epsilon}C_0\varphi^G(d_\mathbf{p}, d_\mathbf{q}) = \frac{|\mathcal{R}|}{2\epsilon}C_0\varphi^G(d_\mathbf{p}, d_\mathbf{q}) \tag{20}$$

Again, if we want the equality to hold, we have to set $C_0 = \frac{2\epsilon}{|\mathcal{R}|}$. For 8 directional SGM this yields $C_0 = \frac{1}{4}$.

We have now shown that, given C_0 and C_1 as above, all tree MRFs are equal for TRW-T ($\widehat{E}$) and SGM (E). The next steps of both algorithms are the same:

1. Peform BP on all trees
2. Sum the beliefs from all scanlines that pass through a pixel (TRW-T equation (5), SGM equation (17))
3. Choose the label that minimizes each belief (TRW-T equation (6), SGM equation (10))

Since the same steps are applied to the same trees, the final labeling is also the same for E and $\widehat{E}$, and thus also for $\widetilde{E}$. □

Summing up, SGM amounts to the first iteration of TRW-T on a MRF with pairwise energies that have been scaled by a constant and known factor.

4 Implications

Belief propagation was one of the earliest techniques for performing inference in graphical models. Comprehensive literature on its properties exists. Given the formal link that we presented in the previous section, one can now draw from this vast repertoire when examining or improving on SGM. We pointed out that SGM can be treated like a first step of TRW-T, which uses BP at its core. We now exploit the lower bound computed by TRW-T to construct a per pixel uncertainty measure for the depth map.

We propose a novel uncertainty measure for SGM that is based on SGM's relation to TRW-T. For this we interpret the difference between sum of local optima (lower bound) and global optimum as scanline disagreement.

Many uncertainty measures have been presented in the past. For instance [2, 12] model uncertainty based on the image acquisition process. Other evaluations focus on measures that only look at the resulting energy distribution by disparity per pixel [6]. While most of the latter examine the cost curve (of cost vs. disparity) around the optimal disparity, they cannot take into account how this curve was computed as they do not assume any knowledge about the optimization technique. The evaluations in [6] show that none of the different measures is a clear favorite.

Kohli on the other hand derives a confidence measure from the min-marginals of an MRF optimization using dynamic graph cuts [8]. His measure $\sigma_{d_\mathbf{p}}$ is given by the ratio of the current max-marginal and the sum of all max-marginals for a certain node. For this ratio, transforming from EMMs to max-marginals is feasible as the partition function cancels out.

$$\sigma_{d_\mathbf{p}} = \frac{\exp(-\beta_p(d_\mathbf{p}))}{\sum_{\widetilde{d_\mathbf{p}} \in \Sigma} \exp(-\beta_p(\widetilde{d_\mathbf{p}}))} \tag{21}$$

In contrast, our proposed measure depends on the *Union Jack* optimal solution given by SGM, and its lower bound found by applying TRW-T. For this we decompose the *Union Jack* into eight *half* scanlines that start at the image border and end in **p**. Following TRW-T's reparametrization for this decomposition yields equal weights for all trees (lines) $\rho^i = \frac{1}{8}$. Every node and edge appears only once, thus $\rho_\mathbf{s} = \rho_\mathbf{st} = \frac{1}{8}$, only the root is replicated in all eight scanlines, $\rho_\mathbf{p} = 1$. The energy per *half* scanline can then be related to the energy $E^i_{SGM}(D)$ computed by SGM as

$$E^i(D) = \sum_{\mathbf{s} \in V^i} \tfrac{1}{\rho_\mathbf{s}} \varphi_\mathbf{s}(d_\mathbf{s}) + \sum_{(\mathbf{s},\mathbf{t}) \in \mathcal{E}^i} \tfrac{1}{\rho_{st}} \varphi_\mathbf{st}(d_\mathbf{s}, d_\mathbf{t}) \tag{22}$$

$$= \varphi_\mathbf{p}(d_\mathbf{p}) + \sum_{\mathbf{s} \in V^i \setminus \mathbf{p}} 8\varphi_\mathbf{s}(d_\mathbf{s}) + \sum_{(\mathbf{s},\mathbf{t}) \in \mathcal{E}^i} 8\varphi_\mathbf{st}(d_\mathbf{s}, d_\mathbf{t}) \tag{23}$$

$$= 8E^i_{SGM}(D) - 7\varphi_\mathbf{p}(d_\mathbf{p}) \tag{24}$$

According to equation (7), the weighted sum of optimal solutions of the subproblems now yields a lower bound to the optimal *Union Jack* energy E_{UJ}:

$$\min_{D \in \Sigma^N} E_{UJ}(D) \geq \sum_{i=1}^{8} \min_{D^i \in \Sigma^N} \rho^i E^i(D^i) = \sum_{i=1}^{8} \min_{D^i \in \Sigma^N} \left(E^i_{SGM}(D^i) - \tfrac{7}{8}\varphi_\mathbf{p}(d^i_\mathbf{p})\right) \tag{25}$$

$$\Rightarrow \min_{d_\mathbf{p} \in \Sigma} \left(S(d_\mathbf{p}) - 7\varphi(d_\mathbf{p})\right) \geq \sum_{i=1}^{8} \min_{d^i_\mathbf{p} \in \Sigma} \left(L^i(d^i_\mathbf{p}) - \tfrac{7}{8}\varphi_\mathbf{p}(d^i_\mathbf{p})\right) \tag{26}$$

An obvious case when this lower bound is tight, is when all directions choose the same label $d_\mathbf{p}$. This allows us to interpret the difference between the left and right hand side of (26) as disagreement between the *half* scanlines that end in **p**. Figure 2 shows a qualitative evaluation of the difference between the left and right hand side for the Tsukuba scene from the Middlebury benchmark [14]. One would expect the disagreement to be large at object borders which impose depth discontinuities, but there are also some other uncertain regions in the background on the cupboard (see the purple frame). There the untextured material causes ambiguous disparities, which is nicely highlighted in the uncertainty heat map.

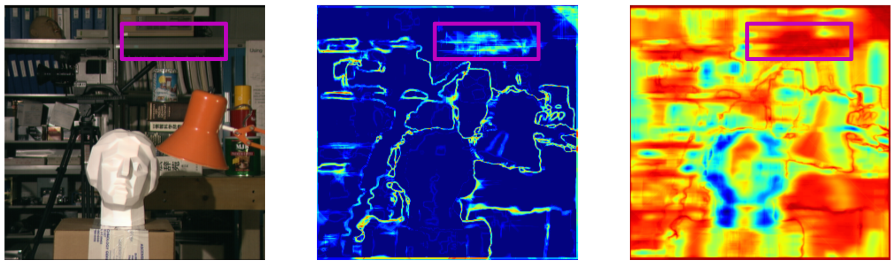


Fig. 2. Left: Cropped Tsukuba image from the Middlebury Benchmark. **Center:** Heat map visualization of our proposed uncertainty measure. **Right:** The uncertainty measure introduced by Kohli [8].

This difference between sum of scanline optima and *Union Jack* optimum can easily be exploited as uncertainty measure. This only incurrs a small overhead of computing the right hand side of (26) by finding the minimum for each direction independently, so searching over $|\Sigma|$ disparities for $|\mathcal{R}|$ directions, and aggregating them. In figure 2 we compare Kohli's uncertainty, computed from *Union Jack* EMMs and transformed to the energy domain $1 - \log(\sigma)$, with our proposed uncertainty measure. The difference between confident and less confident regions is much smoother in Kohli's measure, rendering more parts uncertain, but making it harder to e.g. point a user or algorithm towards regions to investigate. This can be explained by the different interpretations of both measures. A low uncertainty in our measure indicates that we cannot find a much better solution in terms of energy, whereas in Kohli's measure this means that the distribution of max-marginals has a high peak at the current label. It is obvious that for SGM our proposed measure is much easier to interpret.

5 Conclusion

In this work we have derived the formal link between SGM and BP, as well as TRW-T, allowing for a new interpretation of SGM's success and affording new insights. As a specific example, we propose an uncertainty measure for the MAP labeling of an MRF. This characteristic is based on the energy difference between the labeling found and a lower bound, and can be computed efficiently alongside SGM. Such uncertainty measures can be useful for downstream processing of a MAP result. We envision that these insights may encourage the application of SGM as a highly efficient, and now also well-motivated, heuristic to new problems beyond dense stereo matching.

Acknowledgments The research of A.D. and S.A. was partially supported by a Google grant. C.H. and F.A.H. gratefully acknowledge partial financial support by the HGS MathComp Graduate School, the RTG 1653 for probabilistic graphical models and the CellNetworks Excellence Cluster / EcTop.

References

1. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 23(11), 1222–1239 (2001)
2. Frank, M., Plaue, M., Hamprecht, F.A.: Denoising of continuous-wave time-of-flight depth images using confidence measures. Optical Engineering 48(7), 077003–077003 (2009)
3. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 3354–3361. IEEE (2012)
4. Heskes, T., et al.: Stable fixed points of loopy belief propagation are minima of the bethe free energy. Advances in neural information processing systems 15, 359–366 (2003)
5. Hirschmuller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 2, pp. 807–814. IEEE (2005)
6. Hu, X., Mordohai, P.: Evaluation of stereo confidence indoors and outdoors. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. pp. 1466–1473. IEEE (2010)
7. Ishikawa, H.: Exact optimization for markov random fields with convex priors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 25(10), 1333–1336 (2003)
8. Kohli, P., Torr, P.H.: Measuring uncertainty in graph cut solutions–efficiently computing min-marginal energies using dynamic graph cuts. In: Computer Vision–ECCV 2006, pp. 30–43. Springer (2006)
9. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(10), 1568–1583 (2006)
10. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? Pattern Analysis and Machine Intelligence, IEEE Transactions on 26(2), 147–159 (2004)
11. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann (1988)
12. Perrollaz, M., Spalanzani, A., Aubert, D.: Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection. In: Intelligent Vehicles Symposium (IV), 2010 IEEE. pp. 313–318. IEEE (2010)
13. Russell, S.J., Norvig, P., Candy, J.F., Malik, J.M., Edwards, D.D.: Artificial Intelligence: A Modern Approach. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1996)
14. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision 47(1-3), 7–42 (2002)

15. Schraudolph, N.N., Kamenetsky, D.: Efficient exact inference in planar ising models. In: NIPS. pp. 1417–1424 (2008)
16. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for markov random fields with smoothness-based priors. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(6), 1068–1080 (2008)
17. Wainwright, M.J., Jaakkola, T.S., Willsky, A.S.: Map estimation via agreement on trees: message-passing and linear programming. Information Theory, IEEE Transactions on 51(11), 3697–3717 (2005)
18. Yedidia, J.S., Freeman, W.T., Weiss, Y.: Exploring artificial intelligence in the new millennium. chap. Understanding belief propagation and its generalizations, pp. 239–269. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2003), http://dl.acm.org/citation.cfm?id=779343.779352