

Stereo Reconstruction using High Order Likelihood

Ho Yub Jung

Kyoung Mu Lee

Sang Uk Lee

Department of EECS, ASRI, Seoul National University, 151-742, Seoul, Korea

hojub@diehard.snu.ac.kr

kyoungmu@snu.ac.kr

sanguk@ipl.snu.ac.kr

Abstract

Under the popular Bayesian approach, a stereo problem can be formulated by defining likelihood and prior. Likelihoods are often associated with unary terms and priors are defined by pair-wise or higher cliques in Markov random field (MRF). In this paper, likelihood is proposed using higher order cliques. Numerous patch based matching methods such as normalized cross correlation, Laplacian of Gaussian, or census filters are under the naive assumption that a patch's pixels all have same disparities. However, patch-wise cost can be formulated as higher order clique for MRF so that the matching cost is a function of image patch's disparities. A patch obtained from a projected image by disparity map should provide a better match without the blurring effect around disparity discontinuities. Among patch-wise matching costs, census filter approach can be easily reduced to pair-wise cliques. The experimental results on census filter high order likelihood demonstrate the advantages of high order likelihood over independent identically distributed unary model.

1. Introduction

In dense stereo vision problem, likelihood and smoothness prior are essential components in MRF energy minimization framework. Recently, prior modeling made significant advancement with efficient optimization techniques. First order pair-wise smoothness prior is popular because the maximum a posteriori (MAP) is easily approximated through graph-cut and belief propagation [24] [6]. Likewise, second order prior is effectively implemented for stereo problem by combining multiple solutions together using fusion moves [28]. Also, highly connected non-parametric prior was introduced with advantages in high curvature surfaces [23]. Outside of stereo, high ordered priors such as field of expert image prior [18] and N-Potts model [13] were introduced.

Except for conditional random field stereo, where the posterior is estimated directly from pair-wise cliques [20], matching costs are usually aggregated from unary poten-

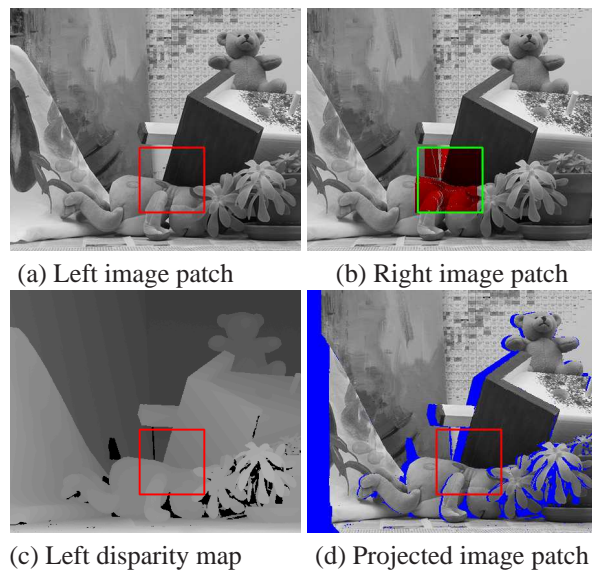


Figure 1. Patch-wise matching assumes that left image patch (a) and right image patch in green (b) are similar, which presumes constant disparities in left patch (a). More plausible match can be found by projecting right image pixels in red (b) by disparity map (c). The projected image patch (d) should provide better match for left image patch (a), except for the occluded areas in blue. The projected image patch (d) is a function of disparity patch (c), which makes the patch-wise matching cost between (a) and (d) to be a high order clique potential.

tials. Basic matching costs are absolute or squared difference under the color consistency assumption. Recent Birchfield Tomasi pixel-wise matching cost provide sampling insensitivities [5]. Plus, there are various window-patch based methods such as zero mean normalized cross correlation, Laplacian of Gaussian (LoG), bilateral background subtraction, rank filter, and census filter [3] [30].

Patch based matching costs are seldom incorporated into MRF framework because of well known fattening effect where objects in front become larger than background. This problem has been addressed by segment based matching [25], variable window matching [26] and adaptive oversegmentation [32]. In [1], disparity discontinuities are pre-

served by assigning penalties for the inconsistencies in the window patches awhile assuming at most two disparities in a patch. The blurring effect can also be minimized by adaptive weight approach based on color and distance differences [29] [8]. These approaches are based on the presumption that similarly colored pixels have similar disparities, which can be interpreted as a restatement of the color regularized smoothness prior.

The basic premise of patch-wise matching approach is that a reference image patch has uniform disparity value regardless of shape and size of a patch. Under the uniform disparity assumption, left and right images are filtered or transformed individually, and unary terms are calculated for each patch. A filter response of reference image patch is independent of disparity map. However, more accurate matching patch can be constructed by projecting the non-reference image with true disparity map. When the difference is taken from a reference patch and projected non-reference patch, the difference becomes a function of disparity values in a patch. Such matching cost is best expressed as high order clique potential in a MRF.

Resulting high order MRF can be optimized using various approaches. High order cliques could be reduced to pair-wise MRFs [10, 2]. Factor node belief propagation can be applied [16]. Markov chain Monte Carlo simulated annealing is most traditional approach [7][31]. Recent cluster based sampling can dramatically increase the efficiency of energy minimization [4]. Also, heuristic window clustering sampler is shown to be more adapt for optimization problems in lattice images [12, 11].

Each of these optimization approaches have different limitations. As clique order becomes higher, the corresponding pair-wise MRF is defined with additional auxiliary nodes and edges. Additional edges and nodes exponentially increase the computational time in belief propagation [16]. Also, general clique order reduction for graph-cut is a significantly time consuming problem in itself [2] [10]. Simulated annealing has heuristic temperature scheduling, inherent randomness, computation time, and weakness in infinite potentials. On the other hand, if MRF is mainly sub-modular and easily be reduced to pair-wise MRF, graph-cut is a practical approach as graph-cut based algorithms have more reliable upper-bound.

The main contribution of this paper is the introduction of general high order matching costs for existing patch-wise matching costs where likelihood is a function disparities in neighboring nodes. Additionally, the high order to pair-wise clique reduction is proposed for census filter matching. Comparison tests against conventional unary matching costs demonstrates how proposed high order matching cost can produce better results, especially around disparity discontinuities.

Next section, we will review MRF stereo and define

some of the notations. Section 3 will introduce general high order matching costs. The pair-wise clique reduction of high order census filter potentials will be discussed. The implementation section explains how prior and occlusion handling are combine with high order likelihood. Also, the details of optimization using QPBOI algorithm will be presented. Experimental section compares high order match, unary patch-wise match and unary pixel-wise match in Middlebury stereo sets.

2. Notations

Given stereo image pair I_L and I_R , obtaining the disparity maps \mathcal{D}_L and \mathcal{D}_R is the goal. Pixel position of an image I will be represent with pixel position set X . $x \in X$. The gray-scale value at pixel position x is denoted as $I(x)$. Rectangular patch is represented as ordered tuple of pixel coordinates $\mathbf{x} = (x_1, x_2, \dots, x_c, \dots, x_{|\mathbf{x}|})$ where x_c is the center pixel of the rectangle. $|\mathbf{x}|$ is the number of elements in a patch \mathbf{x} . $I(\mathbf{x})$ will denote a gray-scaled image patch such that $I(\mathbf{x}) = (I(x_1), I(x_2), \dots, I(x_{|\mathbf{x}|}))$. We will also define a set of patches \mathbb{X} center around each pixel positions such that $\mathbf{x} \in \mathbb{X}$.

The projection function $\pi_{L,R}(x, d)$, projects pixel positions x of I_L to the pixel position at I_R using disparity d . For the rectified stereo pair, the projection functions are simplified to following; $\pi_{L,R}(x, d) = x - [d, 0]$ and $\pi_{R,L}(x, d) = x + [d, 0]$ [28]. For the rest of paper, we will refer I_r as the reference image and I_n as non-reference image instead of I_L and I_R .

The projection function for image patches is formulated as follows.

$$\pi_{r,n}(\mathbf{x}, d) = (\pi_{r,n}(x_1, d), \pi_{r,n}(x_2, d), \dots, \pi_{r,n}(x_{|\mathbf{x}|}, d)). \quad (1)$$

A conventional patch-wise matching minimizes the difference between $I_r(\mathbf{x})$ and $I_n(\pi_{r,n}(\mathbf{x}, d))$. In Fig. 1, the image patch in reference image (a) is $I_r(\mathbf{x})$. The matching patch in (b) is $I_n(\pi_{r,n}(\mathbf{x}, d))$ in green square.

Function f_g is denoted as general patch-wise difference function.

$$\varphi_{\mathbf{x}}(d) = f_g(I_r(\mathbf{x}), I_n(\pi_{r,n}(\mathbf{x}, d))). \quad (2)$$

f_g can be simple as sum of absolute differences or the difference at center pixels. A comprehensive review of various patch-wise matching methods is found in [9].

Likelihood $P(\mathcal{I}|\mathcal{D})$ is assumed to be independent identically distributed (i.i.d.).

$$P(\mathcal{I}|\mathcal{D}) \propto \prod_{\mathbf{x} \in \mathbb{X}} \exp(-\varphi_{\mathbf{x}}(d)). \quad (3)$$

Prior $P(\mathcal{D})$ is a function of the neighboring disparity values.

$$P(\mathcal{D}) \propto \prod_{(x,y) \in \mathbb{N}} \exp(-\varphi_s(d_x, d_y)). \quad (4)$$

\mathbb{N} is set of edges in the neighboring nodes. d_x and d_y are disparity values at their respective pixels. φ_s is smoothness cost. The posterior probability is proportional to the likelihood and prior, $P(\mathcal{D}|\mathcal{I}) \propto P(\mathcal{I}|\mathcal{D})P(\mathcal{D})$. Maximizing the posterior is equivalent to minimizing the energy function.

$$E = \sum_{\mathbf{x} \in \mathbb{X}} \varphi_{\mathbf{x}}(d) + \sum_{(x,y) \in \mathbb{N}} \varphi_s(d_x, d_y). \quad (5)$$

3. High Order Matching Costs

The i.i.d. assumption is convenient because unary costs do not complicate MRF energy minimization. When patch-wise matching costs are used for likelihood, however, i.i.d. assumption is shown to be faulty through fattening effect. We propose stereo likelihoods that are dependent on disparities in the neighboring pixels.

3.1. General High Order Likelihood

Instead of projecting an image patch with single disparity d , a patch is projected with \mathbf{d} .

$$\pi_{r,n}(\mathbf{x}, \mathbf{d}) = (\pi_{r,n}(x_1, d_{x_1}), \dots, \pi_{r,n}(x_{|\mathbf{x}|}, d_{x_{|\mathbf{x}|}})). \quad (6)$$

$\mathbf{d} = (d_{x_1}, d_{x_2}, \dots, d_{x_{|\mathbf{x}|}})$ is a tuple of disparity values at image patch \mathbf{x} . General patch-wise matching cost formulation (2) is slightly modified below.

$$\varphi_{\mathbf{x}}(\mathbf{d}) = f_g(I_r(\mathbf{x}), I_n(\pi_{r,n}(\mathbf{x}, \mathbf{d}))). \quad (7)$$

Thus, $\varphi_{\mathbf{x}}(\mathbf{d})$ is a function of disparity values at image patch \mathbf{x} instead of single disparity value.

In Fig. 1, disparities \mathbf{d} at patch \mathbf{x} are shown in (c) with red square. The projected right image is in Fig. 1 (d). Projected pixels in the right image are marked with red (b). The projected image patch $I_n(\pi_{r,n}(\mathbf{x}, \mathbf{d}))$ should provide a better match than simple window matching technique.

$$P(\mathcal{I}|\mathcal{D}) \propto \prod_{\mathbf{x} \in \mathbb{X}} \exp(-\varphi_{\mathbf{x}}(\mathbf{d})). \quad (8)$$

If the pair-wise prior of equation (4) is maintained, the high order matching energy function is proposed as follows.

$$E = \sum_{\mathbf{x} \in \mathbb{X}} \varphi_{\mathbf{x}}(\mathbf{d}) + \sum_{(x,y) \in \mathbb{N}} \varphi_s(d_x, d_y). \quad (9)$$

3.2. Reduction to Pair-wise Cliques

Depending on the choice of patch-wise matching cost f_g , reduction to pair-wise cliques can be made simple without additional nodes. High order sum of absolute or squared differences can be trivially expressed as sum of unary pixel-wise matching costs. The matching costs that patch-wise filter stereo images has more interesting high order transformation.

Among filtering based methods, the census filter is top ranking non-parametric matching cost in Hirschmuller and Scharstein's matching cost evaluations [9]. When transformed into high order matching cost, it can also be easily reduced to sum of pair-wise functions.

Census matching apply a filtering function $T_c(\cdot)$ on the image patches before finding the patch difference.

$$\varphi_{\mathbf{x}}(\mathbf{d}) = f_g(T_c(I_r(\mathbf{x})), T_c(I_n(\pi_{r,n}(\mathbf{x}, \mathbf{d}))))). \quad (10)$$

Census filter outputs binary bit string, where each bit corresponds to the relative intensities of pixel values around the pixel of interest [30]. The center pixel $I(x_c)$ is evaluated against neighboring pixels $I(x_i)$. If $I(x_i) < I(x_c)$, then i^{th} bit is set to 1, else if $I(x_i) \geq I(x_c)$, i^{th} bit is set to 0. The conventional census filter approach separately transforms both left and right image and computes the Hamming distance between corresponding bit strings.

$$T_c(I(\mathbf{x})) = (C(I(x_c), I(x_1)), C(I(x_c), I(x_2)), \dots)$$

$$C(I(x_c), I(x_i)) = \begin{cases} 1 & \text{if } I(x_c) < I(x_i) \\ 0 & \text{if } I(x_c) \geq I(x_i) \end{cases}. \quad (11)$$

The census filtering function $T_c(\cdot)$ calculates bit string vector from image patch $I(\mathbf{x})$.

Let $T_c(I_r(\mathbf{x})) = \mathbf{b}^r = (b_1^r, b_2^r, \dots, b_{|\mathbf{x}|}^r)$ be the census filter tuple from reference stereo image. The response from the projected image patch is $T_c(I_n(\pi_{r,n}(\mathbf{x}, \mathbf{d}))) = \mathbf{b}^n = (b_1^n, b_2^n, \dots, b_{|\mathbf{x}|}^n)$. The Hamming distance is found by summing absolute distance from each bit.

$$f_g(\mathbf{b}^r, \mathbf{b}^n) = \sum_{i=1}^{|\mathbf{x}|} |b_i^r - b_i^n|. \quad (12)$$

The census filter response b_i^r of reference image is not function of the disparity map and it can be pre-calculated. However, b_i^n is a function of disparities which can be formulated from equation (11).

$$b_i^n = C(I_n(\pi_{r,n}(x_c, d_{x_c})), I_n(\pi_{r,n}(x_i, d_{x_i}))). \quad (13)$$

b_i^n is a function of d_{x_c} and d_{x_i} . d_{x_c} is the disparity value at the center pixel x_c in the patch \mathbf{x} . Likewise, d_{x_i} is the disparity value at pixel i^{th} pixel at \mathbf{x} . By combining Hamming distance equation (12) and the bit response function (13), the high order census matching cost is formulated.

$$\varphi_{\mathbf{x}}(\mathbf{d}) = \sum_{i=1}^{|\mathbf{x}|} |b_i^r - C(I_n(\pi_{r,n}(x_c, d_{x_c})), I_n(\pi_{r,n}(x_i, d_{x_i})))|. \quad (14)$$

The above equation calculates census matching cost from reference image patch and the image patch projected from different image. Equation (14) is a sum of pairwise functions of disparity values at x_c and x_i . MRF, then, becomes pairwise, highly connected, and non-submodular. Such MRF can be minimized effectively using QPBOI- α -expansion algorithm.

4. Implementation

7×7 census filter is used. High order census matching cost defined in equation (14) is expanded to a likelihood.

$$P(I|\mathcal{D}) \propto \prod_{\mathbf{x} \in \mathbb{X}} \prod_{i=1}^{|\mathbf{x}|} \exp(-\varphi_m(d_{x_c}, d_{x_i})). \quad (15)$$

φ_m is defined from equation (14).

$$\varphi_m(d_{x_c}, d_{x_i}) = |b_i^r - C(I_n(\pi_{r,n}(x_c, d_{x_c})), I_n(\pi_{r,n}(x_i, d_{x_i})))|. \quad (16)$$

φ_m will be used from now on to represent high order census matching cost. Following subsection will discuss prior and occlusion handling, and how they can be incorporated into the high order census matching cost.

4.1. Prior

Nonparametric smoothness prior recently introduced in [23] is chosen. The smoothness cost is weighted by color and position differences between the center pixel and neighboring pixels. In the original paper, neighborhoods of 81×81 are used. However, in our experiments, smaller neighborhoods of 7×7 are considered, according to the size of census filter. The smoothness is truncated linear model with weights normalized for each neighborhood set. The pair-wise potentials are summarized below.

$$\varphi_s(d_x, d_y) = w_{x,y} \min(|d_x - d_y|, 2), \quad (17)$$

$$w_{x,y} = \frac{\exp(-\frac{|x-y|}{\sigma_x}) \exp(-\frac{|I(x)-I(y)|}{\sigma_c})}{\sum_{(x,y) \in \mathbb{N}} \exp(-\frac{|x-y|}{\sigma_x}) \exp(-\frac{|I(x)-I(y)|}{\sigma_c})}.$$

In above equation, $I(x)$ and $I(y)$ are color vectors. σ_x and σ_c are associated bandwidths. $\sigma_x = 5$ and $\sigma_c = 10$ are chosen according to the findings in [23]. Following energy function combines proposed likelihood with smoothness prior.

$$E = \sum_{\mathbf{x} \in \mathbb{X}} \sum_{i=1}^{|\mathbf{x}|} \varphi_m(d_{x_c}, d_{x_i}) + \lambda \cdot \sum_{\mathbf{x} \in \mathbb{X}} \sum_{i=1}^{|\mathbf{x}|} \varphi_s(d_{x_c}, d_{x_i}). \quad (18)$$

λ is a regularization term. \mathbb{X} is a set of all possible 7×7 image patches.

4.2. Occlusion Handling

When matching costs are in unary terms, occlusions can be dealt with visibility or one-to-one constraints [14, 15]. Under the one-to-one constraint, matching costs are dependant on disparity values at same scan-line. If pixel is not occluded by other pixels in the same scan-line, the matching cost is assigned otherwise occlusion cost.

For high order matching cost, occlusion handling becomes more complicated. When matching cost is in function of disparity patch, occlusions in different scan-line effect overall matching costs. This increases the clique order in a MRF that is already very highly connected. Thus, instead of building MRF that can handle both occlusions and disparities simultaneously, we alternate finding occluded areas and depths. The occluded areas are shown in blue in Fig. 1 are assumed to be known as \mathcal{O}_r .

There are two different approaches for handling matching costs given occlusion map. First, if x is occluded, the projected pixel is assume to have same pixel value as the reference image pixel. $I_n(\pi_{r,n}(x, d)) = I_r(x)$ if $x \in \mathcal{O}_r$. Second, the matching cost can be found while disregarding occluded pixels. The first approach is based on the color consistency assumption and presumes that the replacement pixels have same color as the occluded pixels in the non-reference image. When there are exposure or lighting differences in stereo pairs, the second approach is more suitable since color consistency is no longer valid. In this paper, we adopt second approach and remove $\varphi_m(d_x, d_y)$ from energy function if $x \in \mathcal{O}_r$ or $y \in \mathcal{O}_r$.

The energy function (18) is re-parameterized so that pairwise potentials in same edge are summed together but occluded matching costs are taken out.

$$E = \sum_{(x,y) \in \mathbb{N}_r} \varphi_t(d_x, d_y). \quad (19)$$

$$\varphi_t(d_x, d_y) = \delta \cdot \varphi_m(d_x, d_y) + \delta \cdot \varphi_m(d_y, d_x) + \lambda \cdot \varphi_s(d_x, d_y) + \lambda \cdot \varphi_s(d_y, d_x).$$

$$\delta = \begin{cases} 1 & \text{if } x, y \notin \mathcal{O}_r \\ 0 & \text{otherwise} \end{cases}.$$

\mathbb{N}_r is set of edges between neighboring nodes (x, y) such that the distance between x and y is less than or equal to $\sqrt{3^2 + 3^2}$. Both prior and likelihood are encoded in the pairwise terms.

Disparities for both frames are found simultaneously by minimizing following energy function.

$$E = \sum_{(x,y) \in \mathbb{N}_L} \varphi_t(d_x, d_y) + \sum_{(x,y) \in \mathbb{N}_R} \varphi_t(d_x, d_y) + \sum_{(x,y) \in \mathbb{N}_{LR}} \varphi_o(d_x, d_y) + \varphi_o(d_y, d_x). \quad (20)$$

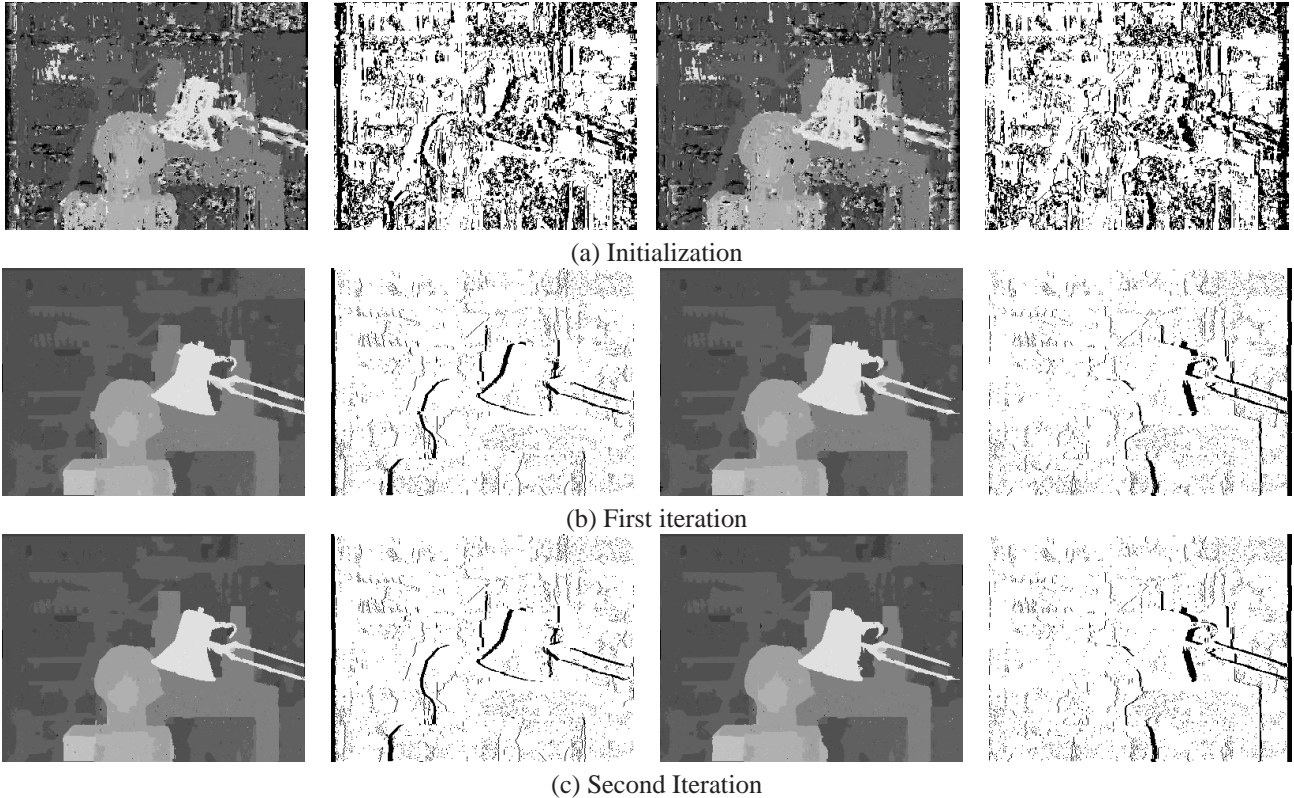


Figure 2. The left image disparity map, left occlusion map, right disparity map and right occlusion map are shown left to right. After initialization using normalized cross correlation (a), we alternate finding disparity maps and occlusion maps. First two iterations are shown in (b) and (c).

\mathbb{N}_L is set of edge when I_L assumes the reference image I_r and I_R assume I_n . Likewise \mathbb{N}_R is the set of edges when $I_R = I_r$ and $I_L = I_n$. The one-to-one constraint cost φ_o is also included in the final energy formulation. \mathbb{N}_{LR} is set of edges between stereo pair along scan-line.

$$\varphi_o(d_x, d_y) = \begin{cases} \lambda_o & \text{if } y = \pi_{r,n}(x, d_x), d_y \neq d_x \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

The initial occlusions are estimated by checking one-to-one correspondence between disparity pair obtained by normalized cross correlation. With occlusion maps for both images, disparities are determined. With disparities, occlusions are found and the process is repeated until convergence or max iteration. We also applied 3×3 median filter as a post process. The algorithm is detailed in Alg. 1. First few iterations are shown in Fig. 2.

4.3. Optimization

QPBOI can optimize non-submodular MRFs and estimates disparities for unlabeled regions [19]. Instead of combining different solutions together, uniform α labeled disparity map is combined with current disparity state using fusion move proposed in [17].

Algorithm 1 Stereo using high order census matching

- 1: Estimate disparity \mathcal{D}_L and \mathcal{D}_R by 5×5 normalized cross correlation.
 - 2: Estimate occlusions \mathcal{O}_L and \mathcal{O}_R for both images by cross checking one-to-one matches in \mathcal{D}_L and \mathcal{D}_R .
 - 3: Estimate disparities \mathcal{D}_L and \mathcal{D}_R by minimizing the energy function (20) using partial QPBOI- α -expansion.
 - 4: If disparities \mathcal{D}_L , \mathcal{D}_R and occlusions \mathcal{O}_L , \mathcal{O}_R are unchanged or if iteration is over max-iteration, terminate. Else, repeat to step 2.
 - 5: Apply 3×3 median filter to \mathcal{D}_L and \mathcal{D}_R .
-

However, high connections are problematic for computation. We could not expand the α label over whole image in single fusion. The high connectivity problem was also addressed in [23] with sparse graph approximation where edges that exhibit minuscule prior are taken out. The sparse graph approximation is efficient when edges are of similar type and dominant edges are apparent. The edges in proposed energy function, however cannot have dominant edges as they are matching costs which should be treated

| | Tsukuba Venus Teddy Cones | | | |
|-----------------------|---------------------------|-------------|-------------|-------------|
| Census high order, TL | 1.03 | 0.11 | 5.49 | 2.39 |
| Census unary, TL | 1.98 | 0.71 | 6.97 | 3.88 |
| Pixel-wise [23], TL | 0.84 | 0.81 | 6.40 | 3.29 |
| High order, CL | 1.30 | 0.22 | 5.62 | 2.40 |
| Unary, CL | 1.98 | 1.32 | 8.49 | 4.15 |
| Pixel-wise [23], CL | 1.12 | 2.23 | 7.25 | 4.46 |

Table 1. Percentage error on non-occluded areas are listed in this table. ‘‘TL’’ stands for tuned regularizing λ value for each individual stereo pair. ‘‘CL’’ stands for constant regularizing term for the test set. Proposed high order matching cost approach is compared against patch-wise unary and pixel-wise matching costs under same smoothness priors. The method with lowest disparity error is in bold.

equally over the whole image. Instead, α -expansion is performed iteratively over random subsets of nodes.

The pixel position $x = (i, j)$ is defined by column index i and row index j . Let us define a set of node $\mathbf{W}_{j,k} \subset X_L \cup X_R$. $\mathbf{W}_{j,k}$ is a set of nodes where row index is less than or equal to j and greater than or equal to k . Partial QPBO- α -expansion Alg. 2 is used to estimate disparities in optimization step 3 of Alg. 1.

Algorithm 2 Partial QPBO- α -expansion

- 1: Set $j = 0$. Randomly select k from set of integers $\{25, \dots, 50\}$.
 - 2: Set alpha label $\alpha = 0$.
 - 3: Hard constraint to current labels for set of nodes that are $X_L \cup X_R \setminus \mathbf{W}_{j,k}$.
 - 4: Update labels for $\mathbf{W}_{j,k}$ using QPBO- α -expansion.
 - 5: Update $\alpha = \alpha + 1$. If α is less or equal to the max disparity value repeat to step 3, else move to step 6.
 - 6: Set j to k . Update k by adding random integer from $\{25, \dots, 50\}$. If k is greater than max row index set k to max row index. If j is equal to max row index, terminate else repeat to step 2.
-

Proposed MRF has full connectivity over 7×7 neighborhoods and additional edges for one-to-one constraint. Partial QPBO- α -expansion iteratively updates subsets of nodes, but only the edges that are connected to $\mathbf{W}_{j,k}$ need to be considered. It reduces computation time and memory. Full α -expansion is desirable, however partial expansion seems to be a good estimation for proposed MRF.

5. Experiments

High order and i.i.d. likelihoods are compared over Middlebury stereo pairs [21, 22, 9]. Same census matching and smoothness prior are applied to MRFs, and they are

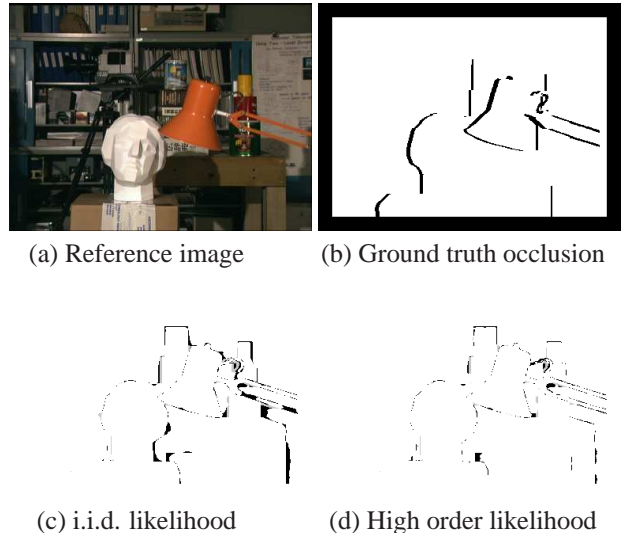


Figure 3. Error areas are shown in black for Tsukuba stereo pairs. 7×7 patch-wise unary matching cost MRF (c) exhibits significant errors around disparity discontinuities. Proposed high order likelihood (d) minimizes the blurring effect around discontinuities awhile keeping 7×7 patch statistical differences as the matching costs.

| | Tsukuba Venus Teddy Cones | | | |
|----------------|---------------------------|------|------|------|
| High order, TL | 5.41 | 1.48 | 14.6 | 6.94 |
| Unary, TL | 10.6 | 8.60 | 18.6 | 10.5 |
| High order, CL | 6.74 | 2.96 | 14.9 | 6.96 |
| Unary, CL | 10.6 | 10.1 | 19.2 | 11.2 |

Table 2. Percentage error around discontinuities are listed in this table. ‘‘TL’’ stands for tuned regularizing λ value for each individual stereo pair. ‘‘CL’’ stands for constant regularizing term.

minimized by QPBO- α -expansion algorithm. The matching cost for unary potentials are formulated in equation (3). 7×7 census filter is applied to stereo pairs separately and simple Hamming distances are found for as unary potentials. Matching costs are updated for each iteration in order to subtract occluded regions out of the matching cost calculations.

All parameters are kept constant except for the smoothness regularizing term λ . We test both for λ optimized for each stereo pair and constant λ for all stereo pairs. Optimal λ s are found separately for unary and high order MRFs. The performances are summarized in Table 1. High order matching cost found lower error than patch-wise unary costs for all 4 Middlebury stereo pairs. The unary potentials’ fattening effect around discontinuities are numbered in Table 2. Additionally, the stereo results from [23] are cited for comparison between pixel-wise matching and high order matching costs. Except for Tsukuba pair, high order matching cost found lower error under same smoothness prior.

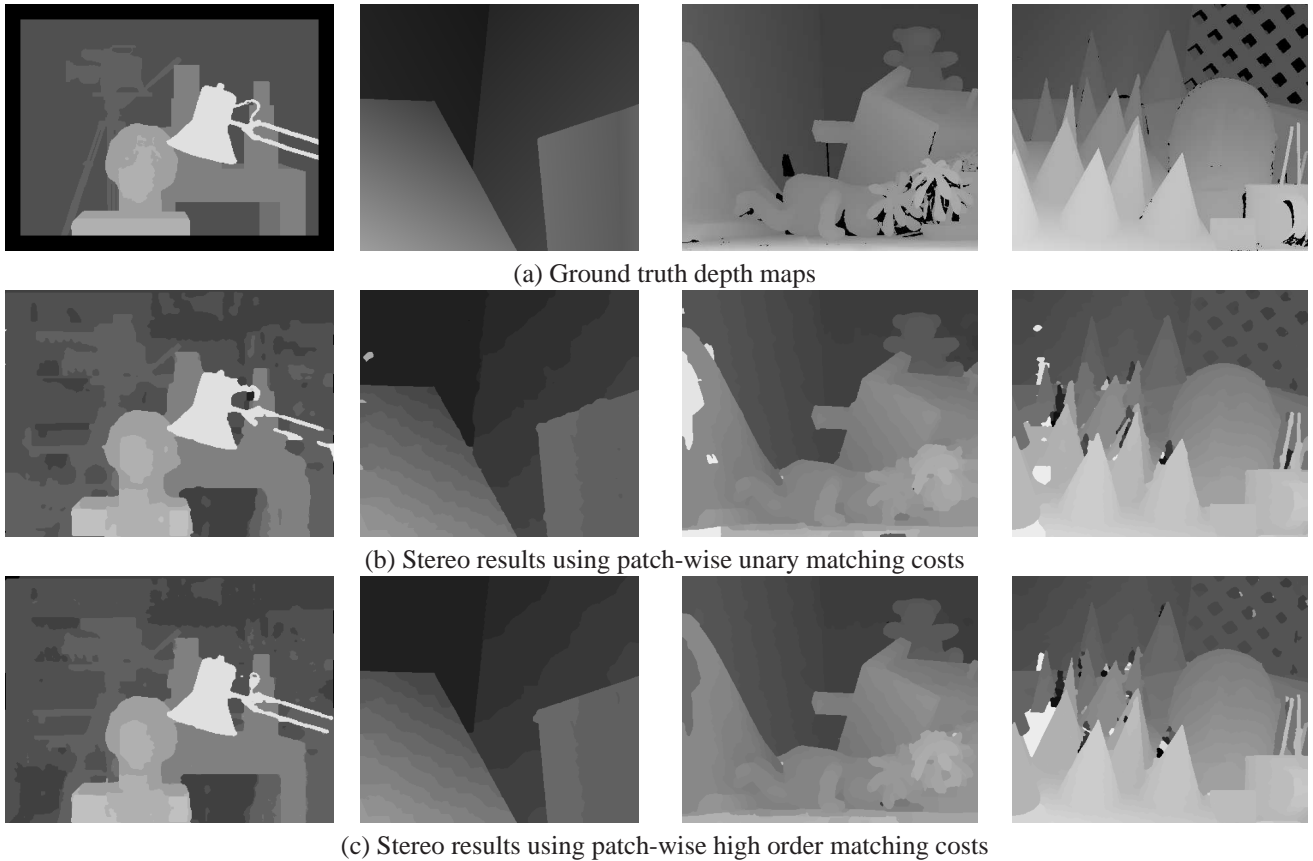


Figure 4. Stereo results from Middlebury pairs using unary and proposed high order matching costs.

Fig. 3 shows the disparity error areas of Tsukuba stereo pair. The color and distant weighted smoothness prior can downsize the blurring effect of the unary potential MRF. Areas where there are minimal color differences have more fattening effect in Fig. 3. Also occluded areas are pruned out from unary matching costs. Thus there are small or no errors around occluded areas for both unary and high order matching MRFs. Regardless, a significant error reduction could be found around disparity discontinuities under high order matching costs. Fig. 4 shows the stereo results for Middlebury stereo pairs.

The computation time for Tsukuba pair is 340 seconds for single iteration of Alg. 2. Venus, Teddy Cones pairs took 371, 1299 and 1200 seconds respectively. The computation time can be an issue but it can be remedied using recent graphics processing unit (GPU) based graph-cut algorithm [27]. Smaller 3×3 or 5×5 high order matching costs are possible with significantly less computation time. However, in our experiments we kept patch size large as possible at 7×7 , in order to demonstrate the differences in the fattening effect more clearly.

6. Conclusions and Future Work

A patch-wise matching cost can measure image patches' statistical differences. Without smoothness prior, a patch-wise matching stereo method generally outperforms a pixel by pixel stereo. However, inherent fattening effect limits the performances of patch-wise matching costs, while smoothness prior allows a discontinuity preserving stereo under pixel-wise matching.

In this paper, we reexamine the assumption behind using patch-wise matching costs as unary potentials. Previous patch-wise matching presume that disparities in a patch are uniform. We proposed matching cost between reference image and a projected non-reference image by disparity map, instead of between two stereo frames. The projected image is a function of disparity map, thus the matching cost is defined only with high order clique.

General high order matching cost MRF can be optimized using various methods. For census filter, high order clique potentials are easily reduced to highly connected pair-wise MRF. The experiments using Middlebury stereo set illustrates the advantages of high order over unary matching costs in eliminating fattening effect. In our future work, the

computation time will be reduced using GPU based Cuda-Cut and more general optimization approaches will be discussed.

References

- [1] M. Agrawal and L. S. Davis. Window-based, discontinuity preserving stereo. *Proc. Conf. Computer Vision and Pattern Recognition*, 2004.
- [2] A. M. Ali, A. A. Farag, and G. L. Gimel'farb. Optimizing binary mrfs with higher order cliques. *Proc. European Conf. Computer Vision*, 2008.
- [3] A. Ansar, A. Castano, and L. Matthies. Enhanced real-time stereo using bilateral filtering. *Int'l Symposium on 3D Data Processing, Visualization and Transmission*, 2004.
- [4] A. Barbu and S.-C. Zhu. Generalizing swendsen-wang cut to sampling arbitrary posterior probabilities. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27, 2005.
- [5] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(5-43), 1998.
- [6] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23, 2001.
- [7] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(721-741), 1984.
- [8] Y. S. Heo, K. M. Lee, , and S. U. Lee. Simultaneous color consistency and depth map estimation for radiometrically varying stereo images. *Int'l Conf. Computer Vision*, 2009.
- [9] H. Hirschmuller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Aug. 2008.
- [10] H. Ishikawa. Higher-order clique reduction in binary graph cut. *Proc. Conf. Computer Vision and Pattern Recognition*, 2009.
- [11] H. Y. Jung, K. M. Lee, and S. U. Lee. Toward global minimum through combined local minima. *Proc. European Conf. Computer Vision*, 2008.
- [12] H. Y. Jung, K. M. Lee, and S. U. Lee. Window annealing over square lattice markov random field. *Proc. European Conf. Computer Vision*, 2008.
- [13] P. Kohli, M. P. Kumar, and P. H. Torr. p^3 and beyond: Solving energies with higher order cliques. *Proc. Conf. Computer Vision and Pattern Recognition*, 2007.
- [14] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusion via graph cuts. *Proc. Int'l Conf. Computer Vision*, 2001.
- [15] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. *Proc. European Conf. Computer Vision*, 2002.
- [16] X. Lan, S. Roth, D. Huttenlocher, and M. J. Black. Efficient belief propagation with learned higher-order markov random fields. *Proc. European Conf. Computer Vision*, 2006.
- [17] V. Lempitsky, C. Rother, and A. Blake. Logcut - efficient graph cut optimization for markov random fields. *Proc. Int'l Conf. Computer Vision*, 2007.
- [18] S. Roth and M. J. Black. Field of experts: A framework for learning image priors. *Proc. Conf. Computer Vision and Pattern Recognition*, 2005.
- [19] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary mrfs via extended roof duality. *Proc. Conf. Computer Vision and Pattern Recognition*, 2007.
- [20] D. Scharstein and C. Pal. Learning conditional random fields for stereo. *Conf. Computer Vision and Pattern Recognition*, 2007.
- [21] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int'l J. Computer Vision*, 2002.
- [22] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. *Proc. Conf. Computer Vision and Pattern Recognition*, 2003.
- [23] B. M. Smith, L. Zhang, and H. Jin. Stereo matching with nonparametric smoothness priors in feature space. *Proc. Conf. Computer Vision and Pattern Recognition*, 2009.
- [24] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo matching using belief propagation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25, 2003.
- [25] H. Tao, H. S. Sawhney, and R. Kumar. A global matching framework for stereo computation. *Int'l J. Computer Vision*, pages 532-539, 2001.
- [26] O. Veksler. Fast variable window for stereo correspondence using integral images. *Proc. Conf. Computer Vision and Pattern Recognition*, 2003.
- [27] V. Vineet and P. J. Narayanan. Cudacuts: Fast graph cuts on the gpu. *Proc. Conf. Computer Vision and Pattern Recognition Workshop*, 2008.
- [28] O. J. Woodford, P. H. S. Torr, I. D. Reid, and A. W. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009.
- [29] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28, April 2006.
- [30] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. *European Conf. Computer Vision*, MAY 1994.
- [31] S. C. Zhu, X. W. Liu, and Y. N. Wu. Exploring texture ensembles by efficient markov chain monte carlo: Toward a trichromacy theory of texture. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(6), 2000.
- [32] C. L. Zitnick and S. B. Kang. Stereo for image-based rendering using image over-segmentation. *Int'l J. Computer Vision*, Special issue 2006.