Cross Image Inference Scheme for Stereo Matching

Xiao Tan[1,2], Changming Sun[2], Xavier Sirault[3], Robert Furbank[3], and Tuan D. Pham[4]

[1] SEIT of UNSW Canberra, Canberra, ACT 2610, Australia
xiao.tan@csiro.au
[2] CSIRO Mathematics, Informatics and Statistics, Locked Bag 17, North Ryde, NSW 1670, Australia
changming.sun@csiro.au
[3] CSIRO Plant Industry, Clunies Ross Street, Canberra, ACT 2601, Australia
{xavier.sirault,robert.furbank}@csiro.au
[4] Aizu Research Cluster for Medical Engineering and Informatics, The University of Aizu, Fukushima 965-8580, Japan
tdpham@u-aizu.ac.jp

Abstract. In this paper, we propose a new interconnected Markov Random Field (MRF) or iMRF model for the stereo matching problem. Comparing with the standard MRF, our model takes into account the consistency between the label of a pixel in one image and the labels of its possible matching points in the other image. Inspired by the turbo decoding scheme, we formulate this consistency by a cross image reference term which is iteratively updated in our matching framework. The proposed iMRF model represents the matching problem better than the standard MRF and gives better results even without using any other information from segmentation prior or occlusion detection. We incorporate segmentation information and the coarse-to-fine scheme into our model to further improve the matching performance.

1 Introduction

Researches have been carried out on stereo matching for many years. To formulate the stereo matching problem, most of the well performed algorithms use the Markov Random Field (MRF) formulation which is based on the assumption that the scene is piecewise smooth. Employing some other information or constraints such as color similarity [1], plane or curved surface hypotheses [2–5], and object recognition [6], stereo matching problem can be solved by minimizing an energy function. In models of all these algorithms, the source image is only used for calculating the correlation or for occlusion detection for the reference image. However, we find that the use of the source image in the stereo matching problem can go further.

The main contribution of this paper lies in the modification of the standard MRF model for stereo matching. Our new model is based on the idea that the

K.M. Lee et al. (Eds.): ACCV 2012, Part IV, LNCS 7727, pp. 217–230, 2013.

labels of matching pixels should be consistent in most parts of both images. Therefore, given the disparity map of the reference image, we can infer the disparity map of the source image and vice versa. Thus, a pixel in the network of our interconnected MRF (iMRF) model is adjacent not only to its neighboring pixels in the same image but also to its potential matching pixels in the other image. As a result, two images are treated as a whole in this model and the labels in both images are updated simultaneously. Quantitative evaluations with ground truths show that by considering the consistency of the potential matching pixels, our new model improves the result from the standard MRF which only considers the consistency of pixels in the neighborhood within one image.

1.1 Previous Work

A survey of stereo matching problems and the quantitative evaluation of disparity estimation algorithms is reported by Scharstein and Szeliski [7]. Stereo matching algorithms can be roughly categorized into local and global algorithms. Local algorithms give acceptable results in the smooth and textured areas with relatively cheaper computation; however, any inappropriate selection of the shape or size of the support windows may cause the incidence of wrong estimation. To solve this problem, many techniques have been proposed using adaptive windows [8, 9], multiple-windows [10], or support weighted windows [11–13].

The global method is characterized by using an MRF stereo formulation which is further converted to the problem of optimization for a specific energy function. The design of an energy function has become the hottest research area in recent years. Employing different constraints such as the uniqueness constraint [14], ordering constraint [15], Ground Control Point constraint [16, 17], and segment constraint [18], these methods regularize the labeling under the Bayes rule. Finding the maximum solution for a specific energy function is usually a NP-hard problem; generally an approximate solution is desired. Several methods such as Mean-Field Annealing [19], Dynamic Programming [20], Graph Cut [21], and Belief Propagation [22], have been proposed to provide the approximate solution to the problem. In the models of most of the approaches mentioned above, only the consistency of labels of neighboring pixels is considered, and the consistency of labels of their matching pixels is ignored.

In previous researches, the labels for the source image are mainly used for occlusion detection. The visibility constraint detects the occluded pixels in the reference image by checking whether there exists at least one matching pixel from the source image [23]. In [24], the labels for the source image are used to define the possible disparity range for a given pixel under the visibility constraint. The unique configuration is used in [25] to enforce each pixel to participate only in one assignment to a pixel in the other image. The cross check requires the labels of two matching pixels to be equal based on the uniqueness constraint [26, 4]. None of these approaches use the labels of the source image when estimating the labels of the unoccluded pixels in the reference image. If one pixel in the reference image matches a pixel in the source image, it is intuitive that the latter pixel has a very high probability to match back to the former one. The

common shortcoming of the above mentioned researches lies on the difficulty of incorporating the hard constraint into the probability inference framework. The method in [27] gives a predefined penalty to the labels which break the label consistency between two views; it may give an over penalty to the horizontally slanted object, as stated in [23].

2 Interconnected MRF Model and Turbo BP Optimization

2.1 Two Properties of Stereo Matching

As disparities of pixels in both images are required in our framework, we consider an image as the reference image when its disparity map is being updated and consider the other image as the source image. Let P and P' be the point sets in the reference and source images, respectively. The set of possible matching pixels of a certain pixel $p\,(p \in P)$ is defined as

$$\psi\,(p) = \left\{p' \in P' | p'_y = p_y, B_l \leq p'_x - p_x \leq B_u\right\} \tag{1}$$

where x and y are the horizontal and vertical coordinates of a pixel. B_l and B_u are the lower and upper bounds of disparity search range. To unify the signs of disparities in both images, we define the disparity between p and p' as $d\,(p, p') = p^l_x - p^r_x$, where p^l is the pixel in the left image of the pixel pair and p^r is the one in the right image.

The first property is called the *equality constraint*: assuming p in the reference image matches p' in the source image with disparity d, if p' also matches p, the disparity of p' is strictly equal to d:

$$d\,(p, p') = d = d(p', p) \tag{2}$$

Another property of the stereo matching problem is that a certain pixel p in the reference image has a one-to-one interconnection to p' in the source image through a given disparity d. That is, there exists only one p' satisfying:

$$p' \in \psi(p) : d(p, p') = d \tag{3}$$

We call this the *interconnection constraint*. These two properties are self-evident considering the definition of disparity. We now describe our iMRF model that applies these two properties into our cross image inference scheme to improve the performance of stereo matching.

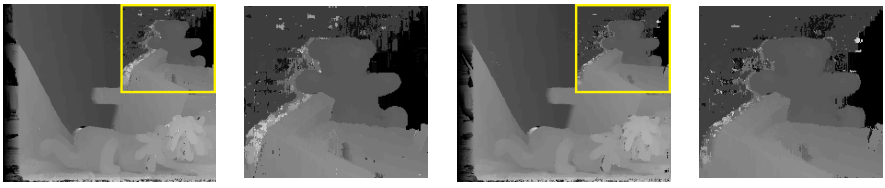
2.2 Interconnected MRF Model

The standard MRF model is used to formulate the local smoothness property in the neighborhood of pixels. However, the dependency between matching pairs is not formulated in this model. In other words, the probability of labeling $d(p, p')$ to p indicates the matching probability between p and p'. On the other hand,

this matching probability also influences the labeling $d(p', p)$ to p'. Denoting by f_p the label of p, we formulate this cross image label dependency as:

$$\Pr\{f_p = d(p, p')\} \propto \Pr\{f_{p'} = d(p', p)\} \tag{4}$$

Hence, given the matching probability of all pixels in one image of the stereo pair, we can obtain the inference for the matching probability of pixels in the other image. As a result, each pixel in the stereo image pair has two matching labels. One corresponds to the MRF model where the pixel is located; the other corresponds to the inference from its possible matching pixels in the other image. Fig. 1 shows a sample of two disparity maps (one is obtained from the data cost of the left image, the other is obtained from the data cost of the right image applying the cross image inference scheme that we proposed). As the possible matching pixels are also in an MRF model, the two MRF models are interconnected. We call this model the iMRF model. In this model, we consider the two labels to be equal for the reason that they are corresponding to the same pixel. The relation between the pixel and its two labels is similar to the relation of the data bit and its interpretations of two code sequences in the turbo coding scheme [28]. Inspired by the implementation of BP to the turbo decoding scheme [29], we give an maximum posterior probability (MAP) estimation to our proposed model using the Max-product BP.



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

Fig. 1. (a) and (b) are the results after applying Winner-Take-All (WTA) matching to the data cost term of the left image. (c) and (d) are the results after applying WTA matching to the cross inference term from the data cost of the right image.

2.3 Turbo BP Optimization

In this section, we first present the network of our model in Fig. 2(a) and then show the message updating rule under the BP framework.

In Fig. 2(a), a solid line between pixels encodes the pair-wise smoothness constraint by a potential function V which can be a Potts model, a linear model or a quadratic model as discussed in [30]. A linear model used in our scheme is:

$$V(f_p, f_q) = \min(\rho_V |f_p - f_q|, T_V) \tag{5}$$

where ρ_V is a parameter which discourages disparity jump between neighboring pixels, and T_V is a truncation threshold which limits the penalty on the disparity jump at the disparity edge of a disparity map. The blue dash lines in Fig. 2(a)

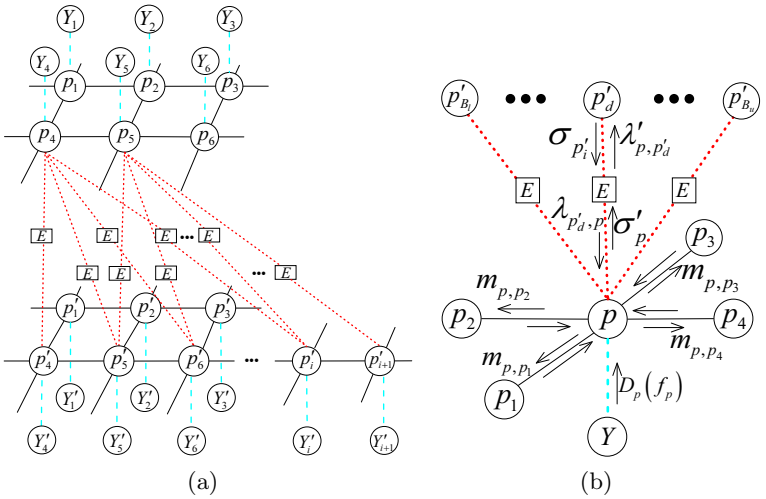


Fig. 2. (a) The network of our iMRF model. (b) A portion of the network showing message exchange.

represent the data costs Y which can be obtained from any correlation calculation algorithm. The red dash lines encode the cross image inference based on the assumption given in Eq. (4). Under this assumption, the labeling probability of a sender node is sent to its receiver node. As discussed in Section 2.1, the labels of two matching pixels should be equal to their disparity under the equality constraint which is denoted by the box with an "E" as shown in Fig. 2(a).

For clarity, a small portion of the full network is shown in Fig. 2(b). Here, we take one image as the reference image. As the two images are treated equally in our model, all discussions in the rest of this section can be similarly applied to the network when the other image is taken as the reference image. Let $N(p)$ be the set of neighboring pixels of p in the same image, p'_d be the matching points of p in the other image with disparity d. We denote a specific neighboring pixel of p in $N(p)$ by q. When the negative log model is used to formulate the probability, the message that p sends to q at iteration t under the Max-product rule is:

$$m^t_{p,q}(f_q) = \min_{f_p} \left(D_p(f_p) + V(f_p, f_q) + \sum_{p'_d \in \psi(p)} \lambda^{t-1}_{p'_d,p}(f_p) + \sum_{p_s \in N(p) \setminus q} m^{t-1}_{p_s,p}(f_p) \right) - \kappa_{p,q} \quad (6)$$

where f_i is the label of node i. D_p is the message sent by the data cost. $\lambda_{p'_d,p}$ is the message sent to pixel p by its possible matching pixel p'_d. $\kappa_{p,q}$ is a normalization factor for preventing overflow, which is constant for f_q but variable for pixel pairs. According to the interconnection constraint, for a given label f_p, only one possible matching pixel whose disparity with p is equal to f_p corresponding with this label. As a result, only one edge in the set of $\lambda_{p'_d,p}$ is activated for a given f_p. We let d be equal to f_p in Eq. (6) and then obtain,

$$m_{p,q}^{t}(f_q) = \min_{f_p} \left(D_p\left(f_p\right) + V(f_p, f_q) + \lambda_{p'_{f_p},p}^{t-1}\left(f_p\right) + \sum_{p_s \in N(p)\setminus q} m_{p_s,p}^{t-1}(f_p) \right) - \kappa_{p,q} \quad (7)$$

The message sent by p to p'_d at iteration t after applying the equality constraint following the Max-product rule is,

$$\lambda_{p,p'_d}^{\prime t}\left(f_{p'_d}\right) = \min_{f_p} \left(\sigma_p^{\prime t}\left(f_p\right) - \log\left(\delta(f_{p'_d} - f_p)\right) \right) = \sigma_p^{\prime t}\left(f_{p'_d}\right) \quad (8)$$

where $\delta()$ is the Dirac function and $\sigma_p^{\prime t}$ is the message sent out by p to p'_d before applying the equality constraint. According to the BP framework and the assumption given in Eq. (4), $\exp\left(-\sigma_p^{\prime t}(f_{p'_d})\right)$ should be proportional to the posterior probability of the label to p, given the labels of its possible matching pixels. The posterior probability is based on the summation over all incoming messages to the node p except the ones from edges between p and its possible matching pixels. We denote the summation of incoming messages for posterior probability calculation as J:

$$J\left(f_p\right) = \sum_{p_s \in N(p)} m_{p_s,p}^{t-1}(f_p) + D_p\left(f_p\right) \quad (9)$$

For different labels of p, the message of cross image inference is sent by different pixels in its possible matching pixel set. In order to make the message proportional to the posteriori probability, a normalization to J over all its possible matching pixels is necessary. For simplicity, we denote componentwise exponentiation and logarithmic on the message x by $x_{\exp}$ and $x_{\log}$. Furthermore, we introduce Pearl's α notation to define an operation on the message, which is similar to the operation on the vector described in [31]. $y = \alpha x$ means that $y\left(i\right) = x\left(i\right)\left(\sum_{k=1}^{n} x\left(k\right)\right)^{-1}$, for $1 \le i \le n$, where n is the dimension of the message. In other words, α converts a message to its probability vector whose elements are proportional to the values in the message. After defining these operations, $\sigma_p^{\prime t}\left(f_{p'_d}\right)$ is given by:

$$\sigma_p^{\prime t}\left(f_{p'_d}\right) = -\left(\alpha(-J)_{\exp}\right)_{\log}\left(f_{p'_d}\right) \quad (10)$$

As discussed in [32, 33], the labeling converges with the increase of the numbers of iterations and the message sent by the cross image inference in the first few iterations is not reliable. So the confidence of $\sigma_p^{\prime t}\left(f_{p'_d}\right)$ should be controlled by the number of iterations:

$$\sigma_p^{\prime t}\left(f_{p'_d}\right) = -w_\sigma\left(\alpha(-J)_{\exp}\right)_{\log}\left(f_{p'_d}\right) \quad (11)$$

where w_σ is a weighting factor which increases with the number of iterations given by $w_\sigma = i/i_{\max}$, where i is the number of iterations that has been performed, $i_{\max}$ is the total number of iterations needed which is a stopping criterion given by users. Since more reliable estimation from the cross image inference will

be obtained in the last few iterations compared with the estimation from the data cost, the effect of data cost on the message passing should be diminished as the number of iterations increases. Therefore, we weight the message from the data cost term by $w_D = 1 - w_\sigma$.

3 iMRF for Stereo Matching

In this section, we introduce our iMRF model into stereo matching via integrating segmentation information and the coarse-to-fine scheme.

3.1 Segmentation Prior

In our iMRF model, we use the segmentation prior which is formulated by 3D planes as a soft constraint. In order to avoid missing the extraction of the correct plane, we extract several possible planes for each segment using current disparity map and weight them accordingly.

We perform sequential RANSAC [34] on the obtained disparity map to calculate plane parameters for N_{R} times of the sequence or until no outliers are left. N_{R} is a parameter controlling the number of planes to be extracted for each segment, which is set to 5 in our implementation. The weight of each possible plane for a segment is given by its average cost. This is based on the fact that a correct plane has a low average cost. Given the cost volume, we define the average cost of an extracted plane as:

$$C^{(j)} = \frac{\sum\limits_{p \in S} D_p\big(f^{(j)}\big)}{\mathrm{card}\,(S)} \tag{12}$$

where j is the index of the plane, S is the set of pixels in a segment, $f^{(j)}$ is the plane-fitted label given by the jth plane. The cost in a stereo matching problem is a discrete function but $f^{(j)}$ is a continuous label. The subpixel estimation is obtained by linear interpolation between two nearest integer labels: $f_-^{(j)}\big(f_-^{(j)} \leq f^{(j)}\big)$ and $f_+^{(j)}\big(f_+^{(j)} \geq f^{(j)}\big)$:

$$D_p\big(f^{(i)}\big) = \big(f_+^{(i)} - f^{(i)}\big) D_p\big(f_-^{(i)}\big) + \big(f^{(i)} - f_-^{(i)}\big) D_p\big(f_+^{(i)}\big) \tag{13}$$

Then we weight the plane by a normalized negative exponent function based on the average cost of the plane:

$$w^{(j)} = \frac{\exp\big(-C^{(j)}\big)}{\sum\limits_{r} \exp\big(C^{(r)}\big)} \tag{14}$$

Given the possible planes and their weights, we use the truncated Total Variance model [22, 35] as our potential function:

$$\rho^{(j)}(f) = -\ln\Big((1 - T_s)\exp\Big(\frac{-\big|f - f^{(j)}\big|}{\eta}\Big) + T_s\Big) \tag{15}$$

where T_s controls the truncation and η controls the penalty of deviation from a fitted result. In this paper, T_s is set to $\exp\left(-\frac{S_r}{10}\right)$ and η is set to 1, where S_r is the disparity search range. Given the potential function and the weightings of planes, we define the plane fitting term as:

$$S\left(f\right) = w_S \sum_j w^{(j)} \rho^{(j)}\left(f\right) \tag{16}$$

where w_S is the weighting of the plane fitting term. Under this constraint, the update rule in Eqs. (7) and (9) should be changed accordingly:

$$m_{p,q}^t(f_q) = \min_{f_p} \Big(D_p\left(f_p\right) + V(f_p, f_q) + \lambda_{p'_{f_p},p}^{t-1}\left(f_p\right) + \sum_{p_s \in N(p)\setminus q} m_{p_s,p}^{t-1}(f_p) + S_p\left(f_p\right) \Big) - \kappa_{p,q} \tag{17}$$

$$J\left(f_p\right) = \sum_{p_s \in N(p)} m_{p_s,p}^{t-1}\left(f_p\right) + D_p\left(f_p\right) + S_p\left(f_p\right) \tag{18}$$

where S_p is the plane fitting term which is defined in Eq. (16) for pixel p.

3.2 Coarse-to-Fine Scheme

There are three types of messages in our algorithm: the smoothness message, the cross image inference message, and the plane fitting message. In our implementation, we use an amended version of Multi-Grid BP [30] to initialize the smoothness message. The initialization of the cross image inference message at a higher level in a pyramidal scheme is obtained by the summation of all messages of pixels in the corresponding block at the finest level, which is calculated using Eq. (18) on the estimation of message m at the finest level and data cost D_p. The plane fitting message is obtained from the plane fitting result, which does not need to be initialized.

We build the cost pyramid from the finest level to the coarsest level as described in [30]. Because a linear cost function is used for the smoothness term, the discontinuity cost is constant. The smoothness message in the next level can be estimated directly from the smoothness message at the corresponding block in the current level. Assuming the finest level is level 1, the cross image inference message and the plane fitting message used at level l is the summation within a block of 2^{l-1} by 2^{l-1} region in the finest level. The estimation is computed using Eqs. (16) and (18). The message m needed in the computation for the cross image inference message and plane fitting is approximated by the message m at the corresponding place in the current level at the latest iteration.

3.3 Procedure for Stereo Matching Using iMRF

The steps of our stereo matching algorithm are:

1. Compute the correlation volume as the data cost; build a cost volume pyramid from the finest level to the coarsest level L. Set current level to L and initialize all messages in the current level to 1.

2. Use the plane fitting scheme described in Section 3.1 on current disparity map for both images and obtain S message as given in Eq. (16).
3. Initialize $i = 1$ and start to update the message.
   (a) Use Eq. (17) to update the message m in both images at level l.
   (b) Use the message m at the current level (l) to approximate the corresponding message m at the finest level. Calculate the estimation of cross image inference message λ at the finest level using Eq. (18) for both images.
   (c) Obtain the cross image inference message λ at the current level (l) for both images by the summation of λ at the finest level.
   (d) $i = i + 1$; if $i = i_{\max}$ go to step (e); otherwise go back to step (a).
   (e) Compute the current disparity map by

$$f_p = \arg\min_f \left(D_p(f) + \sum_{q \in N(p)} m_{q,p}(f) + \lambda_{p'_f,p}(f) + S_p(f) \right) \tag{19}$$

   if $l = 1$ go to step (4); otherwise initialize m at the next level using the corresponding m at the current level (l); initialize λ for next level by the summation of λ at the finest level obtained in step (b); set $l = l - 1$ and then go to step (2).
4. Obtain the disparity map.

In our application, we update the plane fitting term once in each level of the pyramid as computing the plane parameters is relatively expensive.

4 Experimental Results

We implement the algorithm using Visual C++ 2008 and test images from the Middlebury website [36] and our own images. In the first experiment, we do not use the segmentation information to compare our proposed iMRF model with the standard MRF model. Being a generic stereo matching model, our model does not require any specific data cost acquisition scheme. The data cost can be obtained using any different algorithms such as adaptive support-weight approach [13], cross-based approach [37], or 3D-support windows [38]. In our experiments, we use the cross-based approach [37] to calculate the pixel correlation.

Parameters which affect the performance of the two models are ρ_d, T_V, and ρ_V. ρ_d is a value related to the pixel correlation as the data cost. T_V is the penalty to large depth jumps. ρ_V and ρ_d control the smoothness of the result. For simplicity, we set ρ_V to 1 and change ρ_d and T_V in our experiments.

In our experiments, ρ_d is in the range from 0.2 to 0.9 and T_V is set to L_d/N, where L_d is the number of disparity levels, N varies from 5 to 10. Fig. 3 shows the performance of the two models under different parameter settings.

The result shows that the parameter ρ_d has much more influence to the bad pixel percentage than the parameter T_V, which can be seen from Fig. 3. We then test our proposed model on different datasets with different ρ_d and fixed T_V which is set to $L_d/10$. The results and comparisons are shown in Table 1.

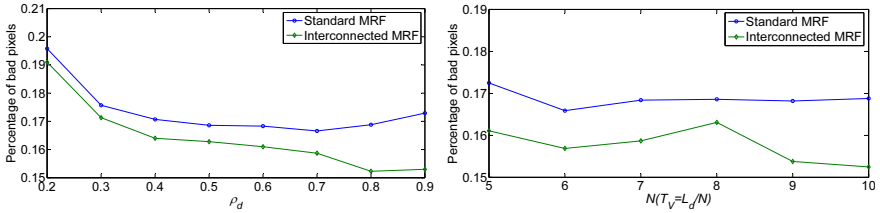


Fig. 3. Left graph: The percentage of bad pixels with respect to ρ_d for the Teddy dataset, by setting $T_V = L_d/10$. Right graph: The percentage of bad pixels with respect to T_V, by setting $\rho_d = 0.8$.

Table 1. Comparison results between the standard MRF and our iMRF according to percentage of bad pixels

ρ_d	0.3		0.5		0.7	
	Standard MRF	iMRF	Standard MRF	iMRF	Standard MRF	iMRF
Teddy	0.176	0.171	0.169	0.163	0.167	0.159
Tsukuba	0.032	0.038	0.028	0.029	0.028	0.027
Venus	0.019	0.018	0.027	0.022	0.026	0.013
Cones	0.111	0.131	0.115	0.114	0.124	0.114

Note: red score represents our iMRF having a better performance.

In our experiments, the iMRF model with our turbo BP algorithm provides a much better performance than the standard MRF with BP optimization. A sample of disparity results corresponding to the last two columns of Table 1 are shown in Fig. 4. Note that, we do not use any other information or constraints such as "segment", "texture", or "occlusion detecting" in our iMRF model. In our experiments, we use the Max-product BP inference scheme for both models with the same parameters. We believe the reason that our model provides better results is due to the fact that the matching problem is better formulated by using the probability inference between two images; however, the standard MRF only considers the consistency between the label of p and the labels of its neighboring pixels within one image and ignores the information from its potential matching points in the other image. In our model, we use the cross images inference term to encourage cross image consistency, which is much closer to the reality of the matching problem.

4.1 Results Using Segmentation

In our next experiment, we test our turbo BP algorithm by incorporating the segmentation information and the occlusion handling scheme as described in Section 3. The parameters used for the rest of the paper are: $\rho_d = 0.7$, $L = 5$, $i_{\max} = 10$, $\rho_V = 1$, $T_V L_d/10$, $w_S = 2 + 0.5\,(L - l)$. L is the number of the levels of the image pyramid, $i_{\max}$ is the number of the iterations within each level.

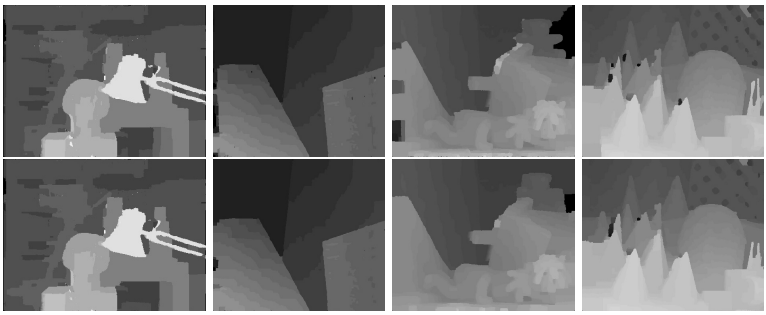


Fig. 4. Top row: Results from the standard MRF model; Bottom row: Results from our iMRF model

We set w_S to $2 + 0.5\,(L - l)$ for two reasons. First, the detailed information may not be available in the disparity map when using coarser level messages to approximate finer level messages; therefore, the plane fitting result is not very reliable for iterations in a coarser level. Second, as one pixel is associated with its four neighboring pixels in the smoothness term, we set the maximum value of w_S to be 4, the maximum discontinuity penalty, with the purpose that a plane fitting result will not break the smoothness constraint.

Our method has the top performance in the algorithms based on the symmetrical model. The comparison is summarized in Table 2. The final and intermediate results together with the ground truths are shown in Fig. 5.

The foreground and background with a slight color difference may be mistakenly regarded as one segment. As a result, some errors may occur at region boundaries due to these possible false segmentations. For example, there is an error at the right part of the paper box in the Tsukuba image pair. The runtime of our algorithm on the Tsukuba dataset without using segmentation information is 45 seconds, and it is 280 seconds when using segmentation information.

The results for some other image pairs in the Middlebury website [36] and our own image pairs are given in Fig. 6. The first two test images in Fig. 6 are from the Middlebury 2006 datasets. The third is a ground view of a tower with textureless sky as the background. The last image is the close-view of a rock. The results show that our algorithm performs well on many different types of images.

Table 2. The results of our algorithm with the Middlebury stereo data and comparisons with other methods which are based on the symmetrical model

Algorithm	Tsukuba			Venus			Teddy			Cones		
	unocc	all	disc	unocc	all	disc	unocc	all	disc	unocc	all	disc
OurMethod	1.14	1.51	5.98	0.17	0.38	2.04	5.72	9.97	15.0	3.14	8.95	8.86
SymBP+occ[23]	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9
Segm+visib[39]	1.30	1.57	6.92	0.79	1.06	6.76	5.00	6.54	12.3	3.72	8.62	10.2
MultiCam GC[24]	1.27	1.99	6.48	2.79	3.13	3.60	12.0	17.6	22.00	4.89	11.8	12.10

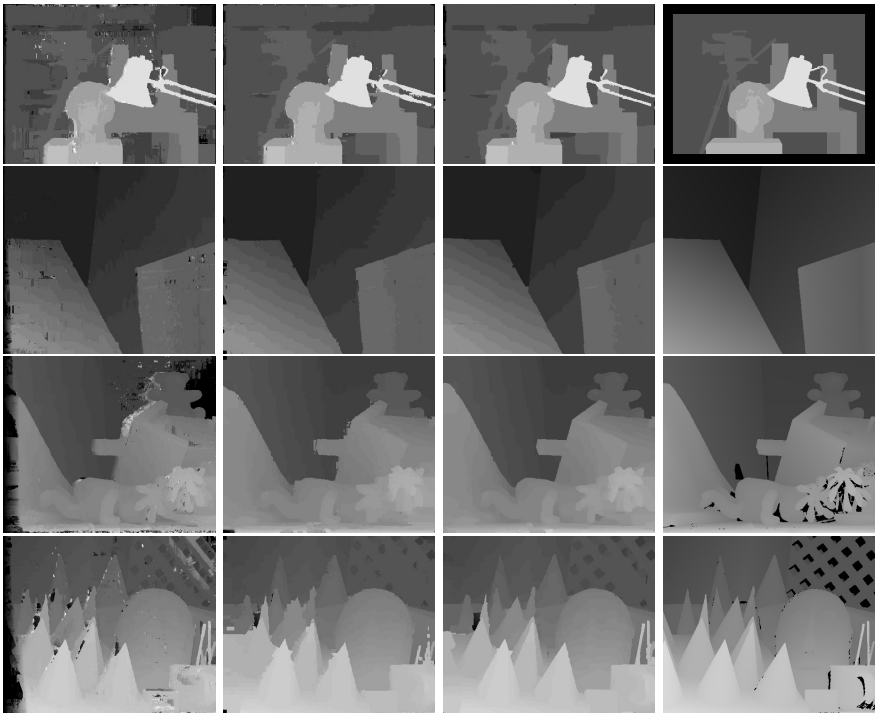


Fig. 5. Fist column: Data costs of datasets; Second column: Intermediate disparity maps from the second level of image pyramid; Third column: Final results of our iMRF based method; Last column: Ground truths of each dataset

Fig. 6. Top row: Original left images; Bottom row: Disparity maps obtained using our iMRF based method

5 Conclusions

A new iMRF model is proposed for stereo matching. In this model, the smoothness term is used for formulating the consistency of the labels of neighboring

pixels. The consistency of matching in two images is formulated by the cross image inference term which is iteratively updated cross both images. We use the Max-product belief propagation on the network of our iMRF model together with the segmentation information and the coarse-to-fine scheme to give an MAP estimation to the disparity problem. Experimental results show that our iMRF model gives a much better estimation to the stereo matching problem than the standard MRF model and the algorithm based on this model provides very good matching results.

Acknowledgements. We thank Chao Zhang for his comments. Tan was partially supported by the China Scholarship Council. Sun was partially supported by the CSIRO's Transformational Biology Capability Platform.

References

1. Larsen, E.S., Mordohai, P., Pollefeys, M., Fuchs, H.: Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. In: ICCV, pp. 1–8 (2007)
2. Birchfield, S., Tomasi, C.: Multiway cut for stereo and motion with slanted surfaces. In: ICCV, vol. 1, pp. 489–495. IEEE (1999)
3. Tao, H., Sawhney, H., Kumar, R.: A global matching framework for stereo computation. In: ICCV 2001, vol. 1, pp. 532–539. IEEE (2001)
4. Yang, Q., Wang, L., Yang, R., Stewénius, H., Nistér, D.: Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. PAMI 31, 492–504 (2008)
5. Bleyer, M., Rother, C., Kohli, P.: Surface stereo with soft segmentation. In: CVPR, pp. 1570–1577 (2010)
6. Bleyer, M., Rother, C., Kohli, P., Scharstein, D., Sinha, S.: Object stereo - joint stereo matching and object segmentation. In: CVPR, pp. 3081–3088 (2011)
7. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV 47, 7–42 (2002)
8. Boykov, Y., Veksler, O., Zabih, R.: A variable window approach to early vision. PAMI 20, 1283–1294 (1998)
9. Veksler, O.: Stereo correspondence with compact windows via minimum ratio cycle. PAMI 24, 1654–1660 (2002)
10. Kang, S.B., Szeliski, R., Chai, J.: Handling occlusions in dense multi-view stereo. In: CVPR, vol. 1, pp. 103–110 (2001)
11. Darrell, T.: A radial cumulative similarity transform for robust image correspondence. In: CVPR, pp. 656–662 (1998)
12. Xu, Y., Wang, D., Feng, T., Shum, H.Y.: Stereo computation using radial adaptive windows. In: ICPR, vol. 3, pp. 595–598 (2002)
13. Yoon, K.J., Kweon, I.S.: Adaptive support-weight approach for correspondence search. PAMI 28, 650–656 (2006)
14. Zitnick, C.L., Kanade, T.: A cooperative algorithm for stereo matching and occlusion detection. PAMI 22, 675–684 (2000)
15. Ishikawa, H., Geiger, D.: Occlusions, Discontinuities, and Epipolar Lines in Stereo. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, p. 232. Springer, Heidelberg (1998)
16. Bobick, A.F., Intille, S.S.: Large occlusion stereo. IJCV 33, 181–200 (1999)

17. Wang, L., Yang, R.: Global stereo matching leveraged by sparse ground control points. In: CVPR, pp. 3033–3040 (2011)
18. Xu, L., Jia, J.: Stereo Matching: An Outlier Confidence Approach. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 775–787. Springer, Heidelberg (2008)
19. Geiger, D., Girosi, F.: Parallel and deterministic algorithms from mrfs: Surface reconstruction. PAMI, 401–412 (1991)
20. Sun, C.: Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. IJCV 47, 99–117 (2002)
21. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. PAMI 23, 1222–1239 (2001)
22. Sun, J., Zheng, N.N., Shum, H.Y.: Stereo matching using belief propagation. PAMI 25, 787–800 (2003)
23. Sun, J., Li, Y., Kang, S.B., Shum, H.Y.: Symmetric stereo matching for occlusion handling. In: CVPR, vol. 2, pp. 399–407 (2005)
24. Kolmogorov, V., Zabih, R.: Multi-camera Scene Reconstruction via Graph Cuts. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 82–96. Springer, Heidelberg (2002)
25. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: ICCV 2001, vol. 2, pp. 508–515. IEEE (2001)
26. Egnal, G., Wildes, R.P.: Detecting binocular half-occlusions: Empirical comparisons of five approaches. PAMI 24, 1127–1133 (2002)
27. Wu, C., Frahm, J., Pollefeys, M.: Repetition-based dense single-view reconstruction. In: CVPR 2011, pp. 3113–3120. IEEE (2011)
28. Berrou, C., Glavieux, A., Thitimajshima, P.: Near Shannon limit error-correcting coding and decoding: Turbo-codes (1). In: IEEE International Conference on Communications, vol. 2, pp. 1064–1070 (1993)
29. McEliece, R.J., MacKay, D.J.C., Cheng, J.F.: Turbo decoding as an instance of Pearl's belief propagation algorithm. IEEE Journal on Selected Areas in Communications 16, 140–152 (1998)
30. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. IJCV 70, 41–54 (2006)
31. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann (1988)
32. Pyndiah, R.M.: Near-optimum decoding of product codes: Block turbo codes. IEEE Transactions on Communications 46, 1003–1010 (1998)
33. Lehmann, F.: Turbo segmentation of textured images. PAMI 33, 16–29 (2010)
34. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24, 381–395 (1981)
35. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60, 259–268 (1992)
36. Scharstein, D., Szeliski, R. (2011), `http://www.vision.middlebury.edu/stereo/`
37. Zhang, K., Lu, J., Lafruit, G.: Cross-based local stereo matching using orthogonal integral images. CSVT 19, 1073–1079 (2009)
38. Richardt, C., Orr, D., Davies, I., Criminisi, A., Dodgson, N.A.: Real-time spatiotemporal stereo matching using the dual-cross-bilateral grid. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part III. LNCS, vol. 6313, pp. 510–523. Springer, Heidelberg (2010)
39. Bleyer, M., Gelautz, M.: A layered stereo algorithm using image segmentation and global visibility constraints. ICIP 5, 2997–3000 (2004)