

Stratified Dense Matching for Stereopsis in Complex Scenes

Jana Kostková and Radim Šára

Center for Machine Perception, Czech Technical University, Prague, Czech Republic
{kostkova,sara}@cmp.felk.cvut.cz, <http://cmp.felk.cvut.cz>

Abstract

Local joint image modeling in stereo matching brings more discriminable and stable matching features. Such features reduce the need for strong prior models (continuity) and thus algorithms that are less prone to false positive artefacts in general complex scenes can be applied. One of the principal quality factors in area-based dense stereo is the matching window shape. As it cannot be selected without having any initial matching hypothesis we propose a stratified matching approach. The window adapts to high-correlation structures in disparity space found in pre-matching which is then followed by final matching. In a rigorous ground-truth experiment we show that Stratified Dense Matching is able to increase matching density $3\times$, matching accuracy $1.8\times$, and occlusion boundary detection $2\times$ as compared to a fixed-size rectangular windows algorithm. Performance on real outdoor complex scenes is also evaluated.

1 Introduction

The core problem of computational stereopsis is computing disparity map of the scene which means (1) finding *correspondences* between binocularly visible points in the input 2-D images and (2) jointly *segmenting* the images to binocularly visible, half-occluded, and mutually occluded regions.¹ There exist two main classes of applications the stereo matching can be used for: view prediction and 3D structure reconstruction. In our research, we are interested in *structure reconstruction* which imposes the following requirements on stereo matching results: it *must not* contain incorrect correspondences and the occluded regions have to be identified accurately. The price paid for these requirements is lower matching density (mainly in texture-less regions). Nevertheless, as our approach belongs to the area-based stereo, the results are required to be as dense as possible.

In general, pixels having the most similar neighbourhoods (measured by various statistics: SSD, SAD, NCC) are assigned as the corresponding pairs. In order to produce accurate results, the matching features have to be as discriminable and stable as possible. By discriminability we mean the ability to recognize correct correspondences. By stability we mean independence on distortions introduced by image projection. Hence, the key-problem is the selection of suitable matching windows over which the statistics are computed. The simplest approaches use centralized fixed-size rectangular windows.

¹To visualize a pair of mutually occluded regions imagine taking a binocular peek through a keyhole. A real example of this phenomenon is shown in the experiments, Sec. 5.

However, due to geometric distortions, this definition fails at curved surfaces and occlusion boundaries. Image similarity computed over independent rectangular image windows has been shown inferior to similarity computed over binocularly corresponding windows.

There exist several approaches trying to cope with this problem. Kanade proposed a method [10], where windows adapt their size: at hypothesized boundaries they are small, otherwise they remain large. Various versions of windows adapting their shape have also been introduced [21, 4, 14, 16]. The resulting non-rectangular windows cover the scene by independent patches (based on the reference image). However, they assume constant disparity within the window. The specific problem at occlusion boundaries is often solved by shifting the window away from the boundary to minimize the non-corresponding background part [2, 8, 15]. Some authors have tried to avoid the occlusion boundary problem by segmentation [20, 22]: The windows are then defined not to cross a segment boundary.

In [16, 5], a matching process starts by finding the seeds (high-probability matches), from which the final solution is spread out. Since the seeds determine the matching accuracy, their correctness is crucial. The first stage we see as a kind of *pre-matching*. We show this can be done in a much more general way. In [3], a *disparity space* has been introduced. First, plausible disparities are computed for each pixel. Based on them, pixels are assigned to *disparity components*, which are defined to be of constant disparity. The final disparity at each pixel is selected by assigning that of its largest component.

We claim that in order to produce good results it is essential to assign the matching over jointly discriminable and stable matching features. By ‘jointly’ we mean that both images contribute to their definition. This seems to be impossible without first having the correspondences. Therefore, we have come to the conclusion, that the matching process should be algorithmically divided into two semi-independent stages, where the first one (pre-matching) hypothesizes reliable matching features, while the second one resolves ambiguities and establishes the final matching. This division follows from requirements on these stages: the pre-matching is to produce *dense results* (low false-negative error) but it need not be a one-to-one mapping. The final matching is to produce *accurate results* (low mismatch and false-positive errors) and it is to be a one-to-one mapping.

In this paper, we follow this view and propose a straightforward method, the *Stratified Dense Matching*. We pose the matching problem in disparity space and design a full-generality pre-matching, which is used to define the windows: They adapt to 3D-connected structures of high similarity, which we call *disparity components*. Such windows are of various shapes, non-constant disparity and are not independent (they are consistent at their overlap). The final matching problem is solved using the similarity statistics re-computed over these adaptive windows.

2 Disparity Components

In our approach, the windows collectively adapt to the 3D-image of the scene in the disparity space. It increases the probability that a high similarity implies a correct match. As a consequence, the discriminability of matching features is improved, which results in higher matching quality. Due to projective distortions this would be impossible to do in input images. The window definition in disparity space enables to get not only windows covering the same scene patches, but also its symmetric left-right and right-left form.

Disparity space T is the set of all possible matches between two or among more images. In the rectified binocular case it can be visualized as matching tables computed separately for each pair of image rows and stacked on top of each other. Each matching

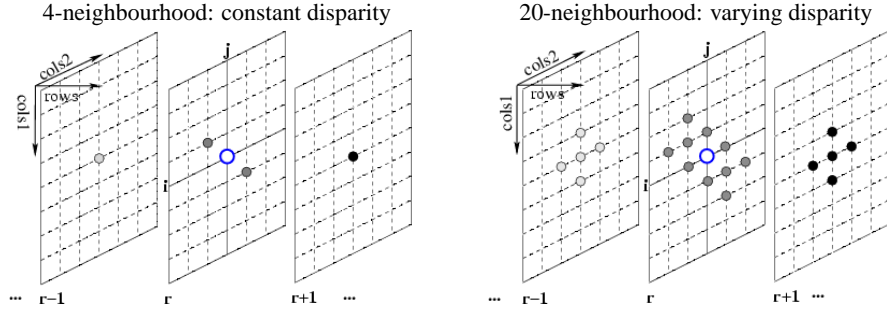


Figure 1: Neighbourhoods of a disparity space point (r, i, j) (empty blue circle): for constant disparity components (left) and for varying disparity components (right).

table consists of similarity values evaluated on the Cartesian product of left and right image pixels in the equivalent row (epipolar line). Matching table elements are called (tentative) pairs. A part of the disparity space (matching tables for rows $r - 1$, r and $r + 1$) is shown in Fig. 1. Matching table rows represent positions i in the left image, the columns represent positions j in the right image. Since the disparity space corresponds to all possible pairwise optical ray intersections it is clear that a surface point neighbourhood in the scene maps on disparity point neighbourhood.

Disparity components are defined in disparity space as connected structures of pairs with high similarity values. The connectedness within a disparity component is defined by a neighbourhood relation. Two high-similarity pairs (r, i, j) and (r', i', j') , where r, r' denote corresponding image rows, i, i' columns in left image and j, j' columns in right image, are neighbours in the disparity space if and only if (1) they are neighbours to each other in the left or the right image, and (2) the difference of their disparities is smaller or equal to a predefined value δ . In mathematical terms the neighbourhood relation can be formulated in the following way:

Definition 1 *The pairs (r, i, j) and (r', i', j') in the disparity space are neighbours to each other if and only if the following three conditions hold:*

1. $|r - r'| \leq 1$,
2. $|i - i'| \leq 1$ or $|j - j'| \leq 1$ for $r = r'$,
 $i = i'$ or $j = j'$ for $r \neq r'$,
3. $|d(r, i, j) - d(r', i', j')| \leq \delta$,

where r, r' are the corresponding image rows, i, i' are the left-image positions, j, j' are the right-image positions and $d(r, i, j) = i - j$ is the disparity of the point (r, i, j) .

Using the neighbourhood relation recursively, the disparity components are traced out. For each pair (r, i, j) , the corresponding disparity component can be identified uniquely (a single match can be part of at most one component). Based on the component, the shape of the adaptive window is found (it brings entire image patches into correspondence).

The parameter δ in the definition of the neighbourhood relation allows disparity variations within one disparity component. If we restrict our definition to points with the same disparity ($\delta = 0$), we get constant disparity components with 4-neighbourhood relation,

see Fig. 1 (left). The 4-connected components correspond to planparallelity assumption, proposed by Boykov [3]. In our approach, the difference of neighbouring pixel disparities is allowed to be smaller or equal to one ($\delta = 1$). Consequently, we get varying-disparity components with 20-neighbourhood relation, see Fig. 1 (right). This definition corresponds to continuity assumption and allows to capture even small variations in disparity.

3 Stratified Dense Matching

In this section we overview the Stratified Dense Matching Algorithm. The input to this algorithm is a pair of left and right rectified images and the output is the disparity map of a scene. The algorithm consists of four steps: (1) *pre-matching*, (2) *disparity component tracing*, (3) *similarity value re-computation*, and (4) *final matching*.

In the first step, an algorithm which is able to produce multivalued disparity maps is run. These *pre-matches* segment the disparity space to a set of connected subsets by eliminating the least number of low-similarity pairs. The requirements on this step are as follows: (1) multivalued disparity map (to obtain all prospective matching hypotheses), and (2) dense results (without unmatched rows or columns in the matching table). The pre-matching step can be based both on global energy minimizations [9, 17, 11] or local correlation methods [10, 18]. We have applied a local correlation method [18].

In the second step, connected disparity components are traced out on pre-matches resulting from the first step. The tracing is based on applying recursively the above-defined 20-neighbourhood relation. For each pre-match, the unique disparity component is identified. Based on the disparity component, the pair of equivalent matching windows is found in the input images. Note that it is not necessary to trace out the entire disparity component, all we need is a set of n -th order neighbours for every pre-match (to get the tracing computationally efficient).

In the third step, for each pre-match, the similarity statistics is *recomputed* using the pair of equivalent matching windows resulting from the second step. In order to get similarity values comparable in their statistical properties, for each pre-match only a fixed-size disparity component neighbourhood is used to jointly define the windows. The match similarity is re-computed only if the corresponding disparity component is large enough (the minimal size is a parameter), otherwise the pre-match is discarded (to suppress mismatches caused by noise or weak textures).

In the fourth step, the final univalued disparity map is computed using the re-computed similarities. The most important requirement is a low error rate. Full density is not strictly required. The results are desired to be as dense as possible, however. In principle, various stereo matching algorithms can be used to compute the final matching. We have selected Confidently Stable Matching [18] for its accuracy: it produces disparity maps that are not necessarily dense but have very low error rates [13].

4 Implementation

Confidently Stable Matching (CSM) solves an optimization task which is defined on mutual occlusion graph $G = (T, E)$ in which the vertex set T is the set of all tentative matches (pairs) and (t, s) is an edge in E if pairs t and s are mutually exclusive, i.e. cannot be both elements of the same matching due to occlusion. We use uniqueness and ordering constraints as the occlusion model. Every pair t in T thus has a set of competitors (neighbours in G) $N(t)$ which we call the inhibition zone of t .² Every pair $t \in T$ has associated the

²Inhibition zone for matchings is as follows: if $t = (i, j)$ then $N(t) = \{(k, l) \mid k = i \text{ or } l = j, (k, l) \neq (i, j)\}$.

value $c(t)$ of match similarity statistics. We say a pair $t \in T$ is confidently excluded by another pair $e \in T$ if $(t, e) \in E$ and $c(t) \leq c(e) - \Delta(t, e)$. The value of $\Delta(t, e)$ is related to the confidence interval widths of $c(t)$ and $c(e)$. Confidently stable subset M is the largest subset of T such that every pair not in M has either no unambiguous competitors in M or is confidently excluded by a competitor in M . Simply and somewhat imprecisely, all pairs not in M are either ambiguous in $c(\cdot)$ or confidently occluded by some strongly better match in M . If exclusion takes into account uniqueness the stable subset is a (univalued) matching but multivalued ‘matchings’ are also possible when the inhibition zone has a finite extent as in Fig. 2. For precise definitions, existence and uniqueness theorems, and the algorithm see [18]. The advantage of CSM is that it does not need a prior detection of matchable image features: it automatically recovers them in the matching process.

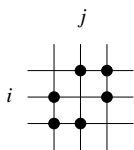


Figure 2: Inhibition zone for pre-matching (black circles) for a pair (i, j) .

To simplify the implementation we use stable matching for the pre-matching step. It is defined as confidently stable matching with $\Delta(\cdot, \cdot) \equiv 0$. The corresponding algorithm is considerably simpler. Multivalued disparity is achieved by using the smallest inhibition zone possible, shown in Fig. 2.

For the component tracing we used a brute-force method: For each pre-match, the neighbours are found by directly testing all their possible positions (based on the defined neighbourhood) without any optimization, which is very time consuming, but we suppose it can be speeded up about 100-fold.

In the third step, we did not recompute the similarities $c(t)$ from image data, we only averaged the values computed in the first step over neighbourhoods in the traced-out disparity component.

There are a few procedure parameters which can be adjusted: disparity search range, initial matching window size, minimal disparity component size, and the confidence level. However, none of them (except for the search range) are critical for the matching process. The default values are: 5×5 initial matching window, and minimal component size of 25 pixels. The confidence level, parametrized by two constants α and β [18], determines the quality of the results. Its choice is left to the user (see the experiments).

On the Middlebury dataset [19] the running time of our current implementation averages 5.2 min, of which pre-matching is 2.1 sec, correlation computation 0.63 sec, final matching 0.64 sec and the rest is spent on component tracing.

5 Experiments

We demonstrate the disparity map improvement due to our adaptive windows over standard rectangular windows. Therefore, in the experiments, we compare the results of Stratified Dense Matching (SDM) with the plain Confidently Stable Matching (CSM) with 5×5 rectangular windows.

We divide the experiments reported here into two groups: the first one is based on a rigorous ground-truth evaluation and focuses on matching failure mechanisms related to insufficient image feature discriminability. The second one demonstrates the results on complex outdoor scenes. For a comparison based on the Middlebury dataset (which is omitted here due to space limitations) see [12].

Ground-Truth Evaluation

This evaluation method [13] is based on a designed artificial scene with known ground-truth, shown in Fig. 3. The scene consists of five thin textured stripes (foreground) in front

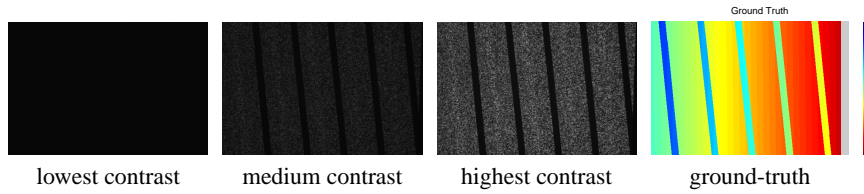


Figure 3: A selection of tested scene: contrasts of 1, 13 and 20, and the ground-truth disparity map. The rightmost bar shows disparity map colour coding: low disparities are dark blue, high disparities are red, half-occluded regions are gray.

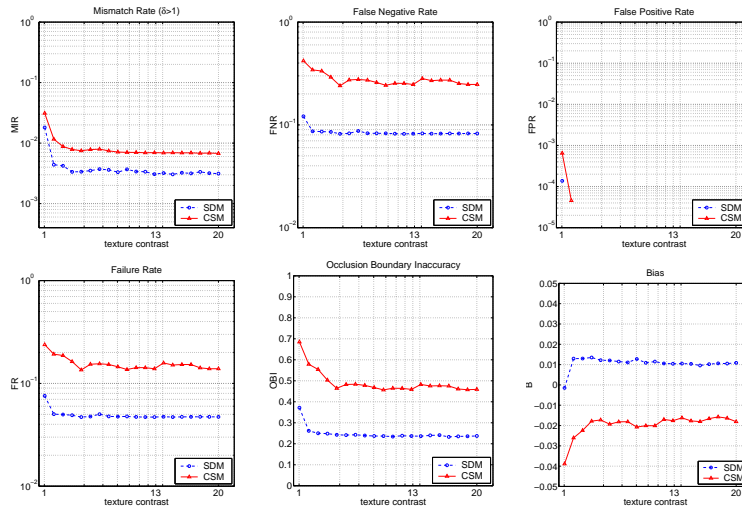


Figure 4: Matching error results. The CSM algorithm: red solid, the SDM algorithm: blue dashed. of a textured plane (background). Twenty stereo images of the scene have been captured under 20 different texture contrast values, which emulate varying signal-to-noise ratio. Three images of those are shown in Fig. 3 (the dark stripes are shadows). The confidence level parameters were set to $\alpha = 20 \sigma^2 / 1000$ and $\beta = 0.05$, where σ^2 is the image intensity variation. The goal of this test set is not to provide a complete cover of all possible stereo data but to expose weaknesses related to image feature discriminability.

Types of Error In the experiment, the following six types of error were distinguished: *Mismatch Rate (MIR)* measures the accuracy of matching, *False Negative Rate (FNR)* measures the disparity map sparsity, *False Positive Rate (FPR)* measures the quality of occluded regions detection, *Failure Rate (FR)* measures the overall disparity map quality (for view prediction), *Occlusion Boundary Inaccuracy (OBI)* measures the precision of occlusion boundary detection, *Bias (B)* measures the algorithm's bias to large objects. The range of all errors is $[0, 1]$, only for bias it is $[-1, 1]$.

Evaluation The results are shown in Fig. 4. Texture contrast (horizontal axis in all plots) is directly related to signal-to-noise ratio. The respective error rates are shown on vertical axes. Note that both axes have logarithmic scale (except in *B* and *OBI* plots).

Both algorithms reach a low level of *MIR* very fast and then they stay constant.³ The *MIR* is normalized by the matching density, we can therefore conclude that the accuracy of both results is better than 99%, while the SDM results improve about $1.8\times$. The False Negative Rate in CSM is about 25%, in SDM about 9%, i.e. the matching density is improved about $3\times$. The *FPR* vanishes in both CSM and SDM except for the worst two contrasts in CSM (where the SDM results are better by about one order of magnitude), which demonstrates the ability of both algorithms to detect occlusions correctly. The *FR* error was improved about $3.3\times$ by applying the SDM approach. Even the *OBI* results are improved: the occlusion boundary is detected about $2\times$ more precisely. None of the algorithms exhibits bias *B* towards large objects. The CSM slightly “prefers” the background (larger object), unlike the SDM, which appears to prefer the foreground (smaller object). However, the bias in the SDM is about $2\times$ smaller than in the CSM.

To conclude, we can say the SDM preserves the good property of CSM (low false positive rate), while the density of the results is improved $3\times$, the occlusion boundary accuracy $2\times$, the matching accuracy $1.8\times$ and the overall error $3.3\times$.

Real Outdoor Scenes

The purpose of this section is to demonstrate the ability of our method to cope even with complex outdoor scenes. We have selected scenes of wide disparity range, thin objects (obstacles) at the foreground, and slanted ground plane. We show our approach is able to correctly detect all the objects in the scene without false positive artefacts (illusions). The confidence level parameters were fixed to the values of $\alpha = 20$ and $\beta = 0.05$ in both SDM and CSM. For comparison, we show the results of the state-of-the-art algorithms: MAP matching via dynamic programming DP (our reimplementation of Cox’s algorithm [7]), and MAP matching via graph cuts GC (the authors’ implementation [11]). The parameters of the two MAP algorithms (the occlusion penalty λ of DP and λ and $penalty_0$ of GC) were manually adjusted to give the visually best results on each of the tested datasets.

The first selected image pair is a photo of a meadow with an apple tree at foreground, a shrub at midrange, and a forest at background. The sky above the scene is completely featureless. The input images together with the results are shown in Fig. 5. The main difference between SDM and CSM is the density, which is about $1.4\times$ greater in SDM (35% of pixels is matched in CSM and 50% in SDM), while the accuracy is preserved. The SDM detects correctly even very fine features in the scene, e.g. the tall grasses at the road side on the right of the tree. The results on the sky and on the tree trunk have been also improved. The DP exhibits the typical “streaking” artefacts around objects (tree crown, shrub), and the tree trunk is disconnected. However, the meadow and the forest are detected correctly. The GC disparity map consists of piecewise constant disparity regions, which do not correspond to the scene at all. The top of the tree crown has been “cut” and the trunk of the tree has not been detected precisely: a higher disparity is assigned to it and this disparity is propagated to the neighbouring meadow, which means GC sees a constant disparity “wall” standing in place of the meadow. A similar frontoparallel wall is hallucinated in place of the meadow on the left of the tree. The forest at the background has not been detected correctly either. Only the sky is correct with unassigned correspondences, but it partially cuts the objects.

The second image is a photo of a larch grove, shown in Fig. 6. In this image pair, we can see a *mutually occluded region*: the area in between the two leftmost trees corresponds

³Which is a remarkable property confirmed in real scenes as well.

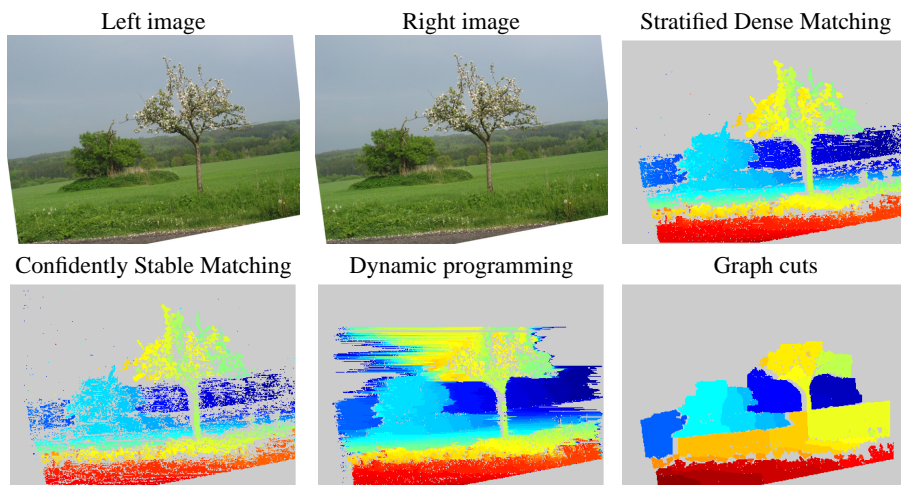


Figure 5: The Apple Tree image pair: the input images are shown together with the results of selected methods. The disparity map colour coding has been described in Fig. 3.

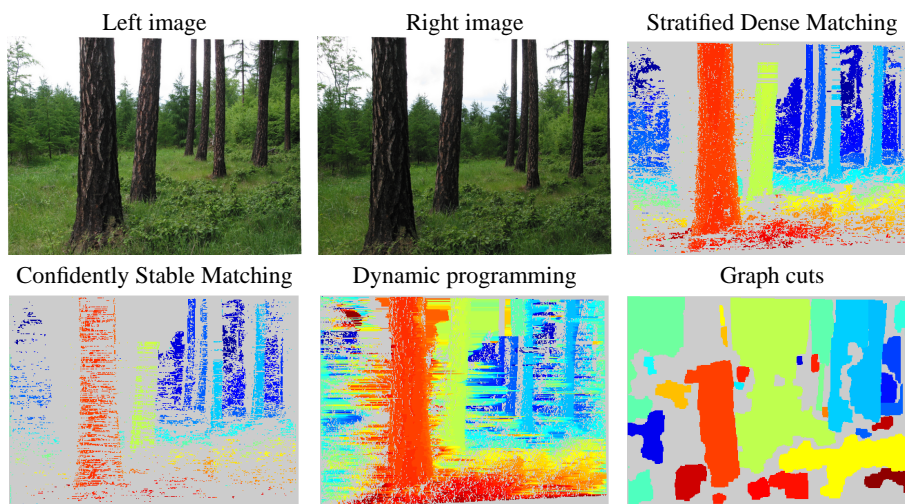


Figure 6: The Larch Grove image pair: the results of the selected methods. The gaps in some of the tree trunks appear due to ordering constraint violation in the scene. For colour coding see Fig. 3.

to different parts of the scene background, thus no correspondences can be found there. This scene violates the ordering constraint, which results in discontinuities in some of the trees. The SDM results are $2.5\times$ denser (CSM matched 19% of all pixels, while SDM 48%), we can find all the trees, ferns on the ground, and the mutually occluded region detected correctly, although there are a few more mismatches there than in the CSM. The DP is not able to cope with occlusions and mutually occluded regions at all. The strong continuity prior causes interpolation artefacts over low-texture (or ambiguous) regions. In GC the parameter tuning was rather difficult. The disparity map shown here has been selected from more than 100 results as visually the best one. The disparity map exhibits

similar artefacts as in the previous scene: Neither mutually occluded nor half-occluded regions have been identified, and many important structural details are missing altogether.

6 Discussion and Conclusions

In this paper, we have proposed a new method, Stratified Dense Matching. Our approach improves the discriminability of matching features by selecting a suitable matching window shape. The windows are defined to adapt to high-correlation structures in disparity space that represent all matching hypotheses. Non-constant disparity within one disparity component is allowed. The window definition in disparity space allows the matching to adapt to slanted and curved surfaces, scale differences, and discontinuities. The stratified matching approach is independent on the selection of the matching algorithm, although it is desirable to use a pre-matching algorithm that guarantees low false negative error and a final matching algorithm that guarantees low false positive error and mismatch error.

We have demonstrated in a quantitative experiment that not only the quality but also the density of disparity map is considerably determined by discriminable joint image features. Further improvement might be achieved by proper local image modeling [6, 1].

Why is the density difference between CSM and SDM so large? The CSM prefers to reject a tentative match if competitors of similar correlation exist. This happens when there are broad correlation value maxima in disparity space due to statistical dependence between neighbouring image pixels, e.g. due to image blur or long correlation length of the visual texture in the scene. In SDM the pre-matching step traces-out the exact positions of the maxima and removes their close competitors. So this step in effect breaks the dependencies by removing the competing matching hypotheses that disturb CSM. The improvement observed in accuracy (*MIR* and *OBI*) results from the adaptation of the matching window shape, which improves the correlation discriminability.

Note that the SDM lies somewhere in between the algorithms that solve one optimization problem per rectified image row (like DP [7]) on the one hand and the algorithms that solve a global optimization problem on the entire disparity space (like GC [11]) on the other hand. The component tracing step is semi-local in its nature but still able to tie together solutions on neighbouring epipolar lines. Its semi-locality brings a great algorithmic advantage over the global optimization approaches.

Our work differs from the work of others in: (1) the definition of disparity components that represent a piecewise contiguous manifold in disparity space and so (2) avoiding the definition of area-based matching over a collection of independent patches, and (3) avoiding explicit local surface model (e.g. planar) by (4) the determination of the window shape directly by the hypothesized solution (as opposed to just selecting the best one from a group of pre-defined windows), (5) in splitting the matching problem into two semi-independent stages with different requirements, (6) re-computing similarity statistics from disparity components, and (7) joining the segmentation of reliable matching features and the correspondence search into a single procedure.

Acknowledgments

This work has been supported by the Grant Agency of the Czech Republic under project GACR 102/01/1371, by the Grant Agency of the Czech Technical University under project CTU 8306413, and by the Czech Ministry of Education under project MSM 212300013.

References

- [1] Stan Birchfield and Carlo Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE PAMI*, 20(4):401–406, 1998.
- [2] Aaron F. Bobick and Stephen S. Intille. Large occlusion stereo. *IJCV*, 33(3):181–200, 1999.
- [3] Yuri Boykov, Olga Veksler, and Ramin Zabih. Disparity component matching for visual correspondence. In *CVPR*, pages 470–475, 1997.
- [4] Yuri Boykov, Olga Veksler, and Ramin Zabih. A variable window approach to early vision. *IEEE PAMI*, 20(12):1283–1294, 1998.
- [5] Qian Chen and Gérard Medioni. A volumetric stereo matching method: Application to image-based modeling. In *CVPR*, pages 29–34, 1999.
- [6] Maureen Clerc. Wavelet-based correlation for stereopsis. In *ECCV*, pages 495–509, 2002.
- [7] Ingemar J. Cox, Sunita L. Higorani, Satish B. Rao, and Bruce M. Maggs. A maximum likelihood stereo algorithm. *CVIU*, 63(3):542–567, 1996.
- [8] Davi Geiger, Bruce Ladendorf, and Alan Yuille. Occlusions and binocular stereo. *IJCV*, 14:211–226, 1995.
- [9] Hiroshi Ishikawa and Davi Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *ECCV*, pages 232–248, 1998.
- [10] Takeo Kanade and Masatoshi Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE PAMI*, 16(9):920–932, 1994.
- [11] Vladimir Kolmogorov and Ramin Zabih. Computing visual correspondence with occlusions using graph cuts. In *ICCV*, pages 508–515, 2001.
- [12] Jana Kostková. Stereoscopic matching: Problems and solutions. RR CTU–CMP–2002–13, Center for Machine Perception, Czech Technical University, 2002.
- [13] Jana Kostková, Jan Čech, and Radim Šára. Dense stereomatching algorithm performance for view prediction and structure reconstruction. In *SCIA*, pages 101–107, 2003.
- [14] R. A. Lane, N. A. Thacker, and N. L. Seed. Stretch-correlation as a real-time alternative to feature-based stereo matching algorithms. *IVC*, 12(4):203–212, 1994.
- [15] Masatoshi Okutomi, Yasuhiro Katayama, and Setsuko Oka. A simple stereo algorithm to recover precise object boundaries and smooth surfaces. In *Proc. of Workshop on Stereo and Multi-Baseline Vision*, pages 158–165, 2001.
- [16] G. P. Otto and T. K. W. Chau. 'Region-growing' algorithm for matching of terrain images. *IVC*, 7(2):83–94, 1989.
- [17] Sébastien Roy. Stereo without epipolar lines: A maximum-flow formulation. *IJCV*, 34(2/3):147–161, 1999.
- [18] Radim Šára. Finding the largest unambiguous component of stereo matching. In *ECCV*, pages 900–914, 2002.
- [19] Daniel Scharstein, Richard Szeliski, and Ramin Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(1):7–42, 2002.
- [20] Hai Tao, Harpreet S. Sawhney, and Rakesh Kumar. A global matching framework for stereo computation. In *ICCV*, pages 532–539, 2001.
- [21] Olga Veksler. Stereo matching by compact windows via minimum ratio cycle. In *ICCV*, pages 540–547, 2001.
- [22] Ye Zhang and Chandra Kambhampettu. Stereo matching with segmentation-based cooperation. In *ECCV*, pages 556–571, 2002.