

# Finding the Largest Unambiguous Component of Stereo Matching

Radim Šára

Center for Machine Perception, Department of Cybernetics  
Faculty of Electrical Engineering, Czech Technical University  
Technická 2, 16627 Prague, Czech Republic  
sara@cmp.felk.cvut.cz

**Abstract.** Stereo matching is an ill-posed problem for at least two principal reasons: (1) because of the random nature of match similarity measure and (2) because of structural ambiguity due to repetitive patterns. Both ambiguities require the problem to be posed in the regularization framework. Continuity is a natural choice for a prior model. But this model may fail in low signal-to-noise ratio regions. The resulting artefacts may then completely spoil the subsequent visual task.

A question arises whether one could (1) find the unambiguous component of matching and, simultaneously, (2) identify the ambiguous component of the solution and then, optionally, (3) regularize the task for the ambiguous component only. Some authors have already taken this view. In this paper we define a new *stability property* which is a condition a set of matches must satisfy to be considered unambiguous at a given confidence level. It turns out that for a given matching problem this set is (1) unique and (2) it is already a matching. We give a fast algorithm that is able to find the largest stable matching. The algorithm is then used to show on real scenes that the unambiguous component is quite dense (10–80%) and error-free (total error rate of 0.3–1.4%), both depending on the confidence level chosen.

## 1 Introduction

The problem of computational stereopsis has been studied in the computer vision field for about four decades [1]. From the very beginning it was clear that binocular stereo matching is an ill-posed problem, since, except in special cases, the solution to a given matching problem is not unique. Not much effort has been spent on understanding the inherent ambiguity of stereopsis, except for the work on the visual hull [2] and the recent work of Baker, Sim, and Kanade [3].

Our working hypothesis is that stereoscopic ambiguity originates from three independent factors: (1) *Data uncertainty* due to insufficient signal-to-noise ratio (SNR) in images of weakly textured objects; (2) *Structural ambiguity* due to the presence of periodic structures combined with the inability of a small number of cameras to capture the entire light field; (3) *Task formulation*: If stereoscopic vision goal is posed as *scene reconstruction* from image projections the problem

is that the same image set can be generated by infinitely many scenes even if the above two factors do not manifest itself [4,2].

The Class 1 ambiguity is the most serious one because it is never completely avoidable. Data uncertainty can be partly reduced by using better cameras or more discriminable match quality measures [5]. It cannot be suppressed by using more cameras [3].

It has long been observed experimentally that more cameras can reduce Class 2 ambiguity significantly [6,7]. To our knowledge the necessary and sufficient conditions for eliminating this ambiguity have not yet been stated.

The Class 3 ambiguity was studied in the work on the photo-hull [2]. Images from any number of cameras are photo-consistent with uncountably many scenes. One avoids the Class 3 ambiguity by separating the stereo matching problem from the scene reconstruction problem. The stereo matching itself, which is nothing but establishing the correspondence mapping between two (or among more) images, is not ill-posed in this sense.

The Class 1 and 2 ambiguities require a prior model to regularize the stereoscopic matching problem. Usually, some form of uniform continuous prior is imposed. The early solutions based on the heuristic of *cohesiveness* [8] culminated in the classical algorithm by Pollard, Mayhew and Frisby [9]. Although some interest remained [10], many authors departed from this path and tried to pose the problem in the framework of statistical estimation theory with a probabilistic continuity prior. The ML estimators of early 90's, based on dynamic programming, were later put on a more sound basis by [11,12] (assuming continuity along epipolar lines) and by others [13,14,15] (assuming isotropic continuity prior). Recent disparity component matching [16] and network flow [17, 18,19] formulations of the matching task also include isotropic prior model but their computational complexity is lower.

Since the most important cause of ambiguity is poor SNR, one has to be very careful in selecting a proper regularizer. Consider a depth discontinuity under low signal-to-noise ratio (from at least one side of the discontinuity image). Strong artefacts may occur if continuity prior is used, see, for instance, the MAP results in Fig. 3, where the thin stripes are not recovered under low SNR. The prior model becomes locally stronger than the constraints given by the input data, which results in interpolation effect within the boundaries surrounding the weakly-textured region. But such interpolation is quite often erroneous, especially in scenes of deep range and relatively thin objects.

Various attempts to switch the regularizer off at discontinuities rely on either detecting image edges under the assumption that they coincide with object boundaries [20,21] or by weighting the regularizer by image gradient [13]. This coincidence does not always occur. Alternatively, they rely on identifying high-tension regions in regularized solution [22], releasing the continuity prior there, and re-computing the solution. Yet another alternative is to use robust regularizers [15]. None of those methods proved fully satisfactory, mainly because they do not work where they are supposed to: under simultaneous occurrence of Class 1 and Class 2 ambiguity.

It seems a possible solution would be to *reject* ambiguous correspondences. If ambiguities of the first two classes are present we may want to pose the problem as finding the unambiguous component of the matching. This has been successfully attempted by Manduchi and Tomasi [23]. They were able to automatically find (Class 2) unambiguous points and match them. The resulting disparity maps are rather sparse. But it is clear that *the aim cannot be at dense matching*. We therefore pose stereo matching as a two-part task:

P1. Establishing the largest possible unambiguous matching.

P2. Partitioning the images into three components:

1. *binocularly visible region* where reliable matches can be found,
2. *monocularly visible region* due to identified occlusions,
3. *ambiguous region* where selecting the unique solution to the matching problem *requires* regularization.

In this paper we will be interested in matching identifying the largest possible image subset as unambiguous at a given confidence level. This is where we differ from [23]. As will be seen in Sec. 4 there are natural scenes where the component is quite large.

Is there any principle that could be used to find the largest unambiguous component of stereoscopic matching in a simple way? Let us first develop a bit of insight on a simple case: suppose (1) there is a signature assigned to each pixel in input images, (2) the signatures of corresponding points are equal, and (3) all pixel signatures are unique. Under such restricted conditions the (actual) matches dominate all alternative matches in their signature similarity measure. It is then easy to find such dominant pairs. We call the result the *dominant matching*. An example of dominant matching algorithm is the popular ‘symmetric maximum method.’ In the noise-free case dominant matching leaves only half-occluded regions unmatched.

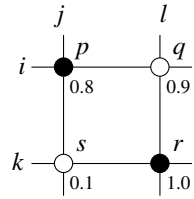
If signatures become *corrupted by random noise* (which is closer to reality), dominant matching fails in a peculiar way. Under Class 1 ambiguity, signature similarity measure can be thought of as a random variable. Some correct matches receive lower similarity and some other candidate matches receive higher similarity. Since there are usually many competitors to a given correct match, the probability of a false match becoming dominant is low compared to the probability of a correct match ceasing to be dominant. The result is that under uncertainty due to noise the dominant matching becomes rather sparse but the component of the matching that survives is exceptionally error-free [24]. The sparsity is due to the strength of the dominance condition. Fortunately, dominant matching turns out to be a special case of stable matching which was introduced in [24]. General stable matching has a greater potential to be dense.

Roughly speaking, matching  $M$  is stable if *every* match  $p \in M$  has no competitor at all or if each of its competitors  $q$  ( $q \notin M$ ) has a competitor  $r \in M$  (cf. Fig. 1). Match  $q$  is a competitor to a match  $p$  if they cannot be both members of the same matching (due to uniqueness and/or ordering constraints, for instance) *and* the signature similarity of  $q$  is greater or equal to that of  $p$  [24].

Obviously, the  $q$  would be a better candidate to be matched if there were no other constraints than the value of its signature similarity alone.

As will be seen later in this paper, stability is a global property of the matched set and stable matching can be computed by a non-iterative algorithm.

**Fig. 1.** Let  $M = \{p, r\}$  be a matching (full circles) on a  $2 \times 2$  matching problem. Match  $p = (i, j)$  of similarity 0.8 has a competitor  $q = (i, l)$  of similarity 0.9. The  $q$  has a competitor  $r = (k, l)$  of similarity 1.0. The pair  $p$  is thus stable in  $M$  (its competitor has a competitor from  $M$ ). The  $r$  is trivially stable since it has no competitor.



After only slight changes in the definition (concerning the sharpness of the inequalities) stable matching is able to reject Class 2 ambiguities in the noise-free case. Unfortunately, it is unable to reject Class 1 ambiguities and Class 2 ambiguities under noise. The reason is two-fold: (1) stable matching as introduced above is always complete,<sup>1</sup> (2) the inequalities in the definition are sensitive to even small noise. In Sec. 3 we formalize the problem of stable matching in a way that there is a margin allowed in the inequalities, which reduces the noise sensitivity. The margin will be computed from confidence intervals for signature similarity measure for each potential pair. We call the result Confidently Stable Matching. It is complete only in the case of no ambiguity in input data and is capable of rejecting both Class 1 and Class 2 ambiguities under noise, which is our ultimate goal. Some experimental results will be shown in Sec. 4.

## 2 Preliminaries

For simplicity we assume throughout this paper that we are solving calibrated stereo problem where epipolar lines coincide with image lines. This simplifies the definition of ordering constraint and makes the definitions simple and intuitive. We will assume matching is done on individual epipolar lines independently. As will be seen in the experimental section this does create any serious artifacts.

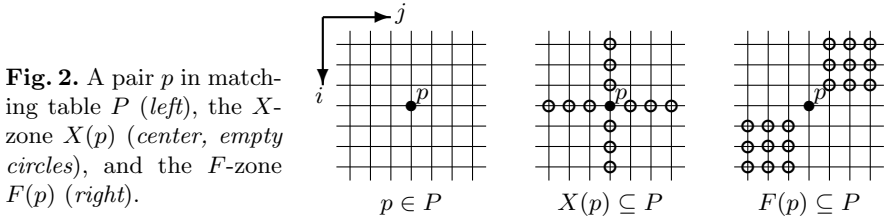
Only binocular matching problem is discussed here as it is the most difficult case in stereoscopic vision. A generalization for polynocular stereo is possible but requires slightly more general theory.

We first need to re-formulate and/or generalize several known concepts so that they better serve our definition of stability and its properties.

Let  $I, J = \{1, 2, \dots, N\}$  be two sets indexing pixels on the left-image and the right-image epipolar lines, respectively and let  $P \subseteq I \times J$ . The element

<sup>1</sup> Complete matching explains *all* pixels as either matched or occluded, see later for a precise definition.

$p = (i, j) \in P$  will be called a (candidate) *pair*. Let  $c$  be the signature similarity measure assigning each pair  $p$  a real number  $c(p)$ .<sup>2</sup> We will visualize  $P$  as a *matching table* as shown in Fig. 2 in which each crossing (circled or not) represents a pair from  $I \times J$ . Note that, in case of sparse matching problems, the  $I \times J$  may or may not be fully populated by  $P$ .



**Fig. 2.** A pair  $p$  in matching table  $P$  (left), the  $X$ -zone  $X(p)$  (center, empty circles), and the  $F$ -zone  $F(p)$  (right).

The  $X$ -zone<sup>3</sup>  $X(p) \subseteq P$  of a pair  $p = (i, j) \in P$  consists of all pairs  $q \in P$  with the same row index  $i$  or column index  $j$ , except for the pair  $p$  itself, as shown in Fig. 2 (center). Similarly, the  $F$ -zone<sup>4</sup>  $F(p) \subseteq P$  of a pair  $p \in P$  is formed by two opposite quadrants around  $(i, j)$  as shown in Fig. 2 (right). The union of  $X(p)$  and  $F(p)$  will be called the  $FX$ -zone,  $FX(p) = X(p) \cup F(p)$ .

A *bipartite matching*  $M$  is a subset of  $P$  in which each  $i \in I$  and each  $j \in J$  is represented at most once. For the sake of brevity, we omit the word ‘bipartite’ in the following text. A subset  $M \subseteq P$  is a matching iff for each  $p, q \in M$  it holds that  $p \notin X(q)$  (this is the so called ‘uniqueness constraint’ [8]). The *cardinality* of matching  $M$  is denoted as  $|M|$ . A maximum cardinality matching has the greatest number of pairs possible.

Matching  $M$  is *monotonic* iff for each two pairs  $p, q \in M, p = (i, j), q = (k, l)$  such that  $k > i$  it holds that  $l > j$ . A subset  $M \subseteq P$  is monotonic iff for each  $p, q \in M$  it holds that  $p \notin F(q)$  (this is the so called ‘ordering constraint’ [26]).

Let us, for a moment, consider the general subsets  $M$  of  $P$  (not necessarily matchings). A pair  $p \in P$  has an *inhibition zone*  $Z(p) \subseteq P$  if it holds that if  $p \in M$  then no pair  $q \in Z(p)$  must be in  $M$ . Two pairs  $p, q$  such that  $p \in Z(q)$  will be called discordant. For instance,  $X$ -zone is the inhibition zone for matchings and  $F$ -zone is the inhibition zone for monotonic subsets of  $P$ . See also Observation 1 later.

Let  $p \in Z(q)$  iff  $q \in Z(p)$ . Then the discordance relation is *symmetric*. In this paper we will assume that inhibition zone  $Z$  always generates symmetric discordance relation. For short we will say the zone is symmetric. The  $X$ -zone and the  $F$ -zone shown in Fig. 2 are both symmetric.

<sup>2</sup> The theory is independent on the choice of this measure, you may safely imagine we talk about normalized cross-correlation computed over small pixel neighborhoods.

<sup>3</sup> The name of the  $X$ -zone was chosen because of its shape.

<sup>4</sup> The  $F$ -zone stands for *forbidden zone* [25].

In the next section we define the confident stability condition, describe its properties in the view of the discussion given in the introduction, and present a simple and fast matching algorithm.

### 3 Confidently Stable Matching

**Definition 1 (Matching Problem).** Let  $P$  be a matching table, let  $c: P \rightarrow \mathbb{R}$  and  $\Delta: P \rightarrow \mathbb{R}^{+0}$  be bounded functions such that  $[c(p) - \Delta(p), c(p)]$  is the confidence interval associated with signature similarity for pair  $p \in P$ , and let  $Z(p)$  be symmetric inhibition zone for each  $p \in P$ . Then  $(P, c, \Delta, Z)$  will be called a matching problem.

We will be interested in matchings that remain stable for the worst possible combination of match similarities from the confidence intervals. Such matchings will certainly be very little sensitive to Class 1 or Class 2 ambiguity. They can be regarded as unambiguous with certainty related to the confidence level chosen. In practice, this is of course limited by the capture probability of the confidence interval estimator.

We will now state the confident stability condition which defines when a subset of matching table  $P$  is considered unambiguous. Later we prove an algorithm that is able to find the largest subset of this property.

**Definition 2 (Confident Stability).** Let  $(P, c, \Delta, Z)$  be a matching problem. A subset  $S \subseteq P$  is confidently stable if for each  $p \in S$  and every  $q \in Z(p)$  the following holds: If  $c(q) \geq c(p) - \Delta(p)$  then there must be  $r \in S \cap Z(q)$  such that  $c(r) > c(q) + \Delta(r)$ .

Thus, stability is a global property of subsets of  $P$ . It is not a one-step look-ahead rule in a discrete minimization strategy. In fact, there is no obvious cost function associated with this problem. Later we will see there is a non-iterative algorithm that finds stable subset of maximum cardinality. See [24] for reasoning that results in this definition. It is easy to see [24] that the following holds:

**Observation 1** If  $X \subseteq Z$  then  $S$  is a matching. If  $F \subseteq Z$  then  $S$  is monotonic.

Before we review the basic properties of Def. 2 we need to introduce the following concept:

**Definition 3 (Completeness).** A subset  $M \subseteq P$  is complete iff for each pair  $p \notin M$  it holds that the intersection  $Z(p) \cap M$  is non-empty.

A complete subset (matching) is ‘the most dense.’ It explains all pixels in an image as either matched or half-occluded. As discussed above, we are not searching for complete matching unless the matching problem is unambiguous. The following lemma tells us what happens if there is no ambiguity:

**Lemma 1.** Let  $(P, c, \Delta, Z)$  be a matching problem. If  $\Delta(p) \equiv 0$  for all pairs  $p$  in confidently stable subset  $S \subseteq P$  and no two discordant pairs are of equal cost  $c(\cdot)$  then  $S$  is complete.

In general, the converse is not true ( $S$  can be complete even if  $\Delta \neq 0$ ). The proof of this lemma is given in [24]. Finally, we give the existence theorem:

**Theorem 1.** *For any given matching problem  $(P, c, \Delta, Z)$  there is a unique (possibly empty) confidently stable subset  $S \subseteq P$  of maximum cardinality.*

The theorem is proven by defining quasi-stable subset  $M \subseteq P$  which is not unique but always exists and is always complete and by showing that any confidently stable subset  $S$  must be a subset of some quasi-stable subset  $M$ . Quasi-stability is defined by choosing  $\Delta \equiv 0$  and making the second inequality not sharp in Def. 2. Then the observation is made that any two complete quasi-stable subsets  $M$  differ in equal-cost discordant pairs only. But such pairs cannot be in  $S$ , which means  $S$  is a subset of the intersection of all quasi-stable subsets of  $P$ . It is then shown that if there are two confidently stable subsets in  $P$  they must be nested. By induction one gets there is a unique confidently stable matching of maximum cardinality, which completes the proof. Details of the proof are found in [27].

It may seem that confidently stable subset is empty quite often. In [5] we show that the sparsity of stable matching is governed by discriminability  $D(c)$  of the signature similarity function  $c$ . Let  $T$  be the set of correct matches for a matching problem. Discriminability is the probability of an inequality event

$$D(c) = P(c(p \mid p \in T) > c(p \mid p \notin T)). \tag{1}$$

As long as  $D(c) > 0.5$  the stable subset is not empty. As shown in a numerical experiment in [5], common similarity measures do differ in their discriminability, but all of them approach the undecidability limit of  $D(c) = 0.5$  for signal-to-noise ratios well below  $-20\text{dB}$ . This means that if (1) images do not exhibit Class 2 ambiguity (repetitive texture) everywhere and (2) we are able to obtain sufficiently tight confidence interval estimates for our image measurements, confidently stable matching will be non-empty.

The algorithm for finding confidently stable matching proceeds as follows:

**Algorithm 1**

**Input:** *Confidently stable matching problem  $(P, c, \Delta, Z)$ .*

**Output:** *Confidently stable subset  $S$ .*

**Procedure:**

1. Initialize  $U := \emptyset, M := \emptyset, S := \emptyset$ , for all  $p \in P$  let  $\tilde{c}(p) := c(p)$ .
2. If  $P$  is empty, terminate.
3. Remove from  $P$  the subset  $L$  of pairs of the largest cost  $\tilde{c}(p)$  in  $P$ .
4. For all  $r \in L$  do:
  - a) if  $r \in U$  then  $R := R \cup Z(r)$
  - b) else if  $r \notin M$  then
    - i.  $\tilde{c}(r) := c(r) - \Delta(r)$ ,
    - ii.  $P := P \cup \{r\}$ ,
    - iii.  $M := M \cup \{r\}$ ,
    - iv.  $U := U \cup Z(r)$ .

5. For all  $r \in (L \cap M) \setminus R$  do:
  - a)  $S := S \cup \{r\}$ ,
  - b)  $P := P \setminus (Z(r) \cup \{r\})$ .
6. Go to Step 2.

**Theorem 2 (Alg. 1 Correctness).** *For any matching problem  $(P, c, \Delta, Z)$  Alg. 1 terminates and finds the confidently stable matching of maximum cardinality.*

In the proof one first notices the difference between confidently stable matching  $S$  and quasi-stable matching  $M$  is that  $S$  contains no pairs in inhibition zones of *converting pairs*  $C \subseteq P \setminus M$ . The task of the converting pairs is thus to remove some pairs from  $M$  to obtain  $S$ . One then shows there is a strict partial ordering on  $C$  for any  $\Delta$  that allows us to uniquely identify  $C$  in  $P$ . One shows that the set  $M$  in the algorithm is quasi-stable matching on  $(P, c, \Delta, Z)$ . The statement of the theorem follows as a corollary. Details are found in [27].

To summarize the algorithm, it constructs a quasi-stable subset  $M$  from pairs in  $P$  in the order of their decreasing cost  $c$  and it simultaneously identifies converting pairs  $C$ , also in the order of their decreasing cost  $c - \Delta$ . The quasi-stable subset is found by the successive selection of the largest-cost pair  $r$  in  $P$  (which is always quasi-stable [24]) in Steps 3 and 4(b)iii followed by the removal of its inhibition zone  $Z(r)$  from the set of all potential matches  $P$  in Step 5b. This is repeated until  $P$  is empty. But every quasi-stable pair  $p$  is forced to cycle through the loop 3–6 twice. When it appears in Step 4 for the first time, its cost  $c(p)$  has the value of the upper limit of the confidence interval. Such pair must be quasi-stable. Its cost is decreased to  $c(p) - \Delta(p)$  and it is returned to  $P$  in Step 4(b)ii. This is followed by the update in Step 4(b)iv of the set  $U$  of all potential converting pairs. The same quasi-stable  $p$  is bound to appear in Step 4 for the second time. Then, unless it is in the inhibition zone of some converting pair  $q \in C$  (i.e. unless it is absent in the set of removed pairs  $R$ ), it is confidently stable and is added to the set  $S$  in Step 5a. Pairs that are not quasi-stable cycle only once; when they are recognized as such by testing their presence in the set  $U$  in Step 4a, it is known they must be converting pairs. A converting pair  $q$  prevents any pair from its inhibition zone  $Z(q)$  to become confidently stable (by adding  $Z(q)$  to the set  $R$ ). Note that some pairs from  $P$  are *never* visited in Step 3, as they are deleted from  $P$  in Step 5b. The correctness of the whole procedure is guaranteed by Theorem 2. Its proof gives more insight into how the algorithm works [27].

The worst-case time complexity of this algorithm is  $O(d^2n)$  if  $P$  is a  $n \times n$  matching table and disparity search interval is  $d$ . The implementation of this algorithm is quite simple for  $FX$ -stable matching. One uses a property of  $FX$ -zones: the union of their complements in  $P$  is a set  $E \subset P$  such that it is blockwise monotonic. The update of the type  $Q := Q \cup FX(r)$  can be done in  $O(d)$  time and the query of the type  $r \in Q$  can be done in  $O(1)$  time. More implementation details are found in [27].



## 4 Experimental Results

We restrict our experimental results to  $FX$ -stable subsets for which  $Z(p) = F(p) \cup X(p)$ . Note that this choice implies the confidently stable subset is a monotonic matching (it obeys ordering constraint). Ordering constraint, represented by the  $F$  part of inhibition zone (see Fig. 2) is a weak regularizer.

We first show results on a designed artificial scene, see Fig. 3. The scene consists of three long thin textured stripes in front of a textured plane. The stripes are approximately 14 pixels wide in  $300 \times 337$  images. The object-background distance was adjusted such that the width of the half-occluded region is exactly equal to the width of the stripes. The scene was illuminated by controlled stabilized illuminant, whose adjustable intensity was used to vary texture contrast. Disparity slowly varies across both the object and the background in the range of  $(-17, 20)$ . Disparity search was done in the full range of  $\pm 337$ . We use this dataset for measuring accuracy of stereo algorithms [24].

For match similarity, modified normalized cross-correlation [28]

$$c(W_L, W_R) = \frac{2 \operatorname{cov}(W_L, W_R)}{\operatorname{var} W_L + \operatorname{var} W_R} \in [-1, 1] \quad (2)$$

was computed for  $5 \times 5$  image windows  $W_L, W_R$ .

A rigorous approach to constructing confidence intervals for stereoscopic matching is discussed in [29]. We approximate the interval by sensitivity of  $c(\cdot, \cdot)$  to independent noise as

$$\Delta(W_L, W_R) = \max \left( \alpha \frac{4 |c(W_L, W_R)|}{\operatorname{var} W_L + \operatorname{var} W_R}, \beta \right), \quad \alpha, \beta \geq 0. \quad (3)$$

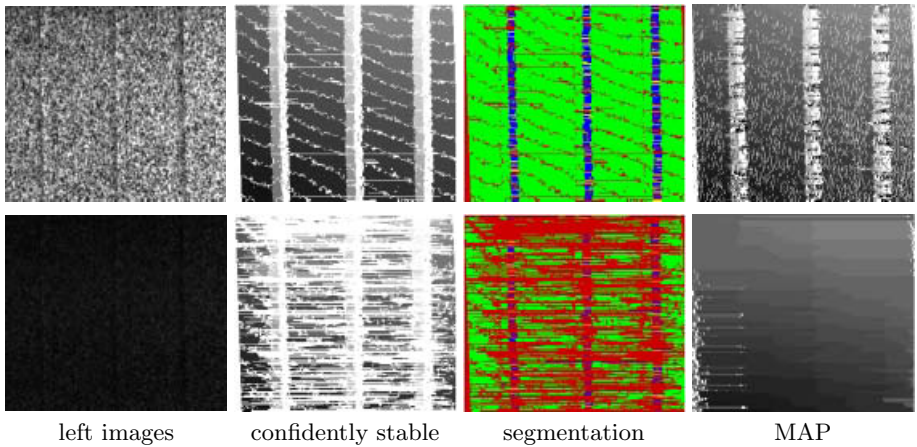
Note it is not necessary that  $\Delta$  measure the image autocorrelation function properties along the epipolar line direction: the  $X$  component of the inhibition zone together with the confident stability constraint takes care of (locally) flat or periodic autocorrelation function.

Computed disparity maps are shown in Fig. 3. White encodes missing matches (half-occlusions and false negatives). The electronic version of this paper uses color-coded maps where errors stand out very clearly.

Low-contrast and high-contrast images of the same scene were tested and compared with ground-truth. The percentage of matching error relative to the matched set is 0.72% in the low-contrast image set and 0.47% in the high-contrast set. Results are compared with the Maximum A Posteriori Probability (MAP) solution using the Cox et al. algorithm based on dynamic programming [30].<sup>5</sup> The matching error percentages for MAP are 59% and 12%, respectively. Errors were evaluated using methodology described in [24] which measures the consistency of matching results over a wide range of SNR.

We demonstrate the performance of Alg. 1 for various choices of  $\alpha$  and  $\beta$  on the Shrub image pair in Fig. 4. The central object of the scene is a sign posted on

<sup>5</sup> Regularization penalty of the Cox et al. algorithm was set to 500 to achieve the best possible results for the high-contrast image pair.

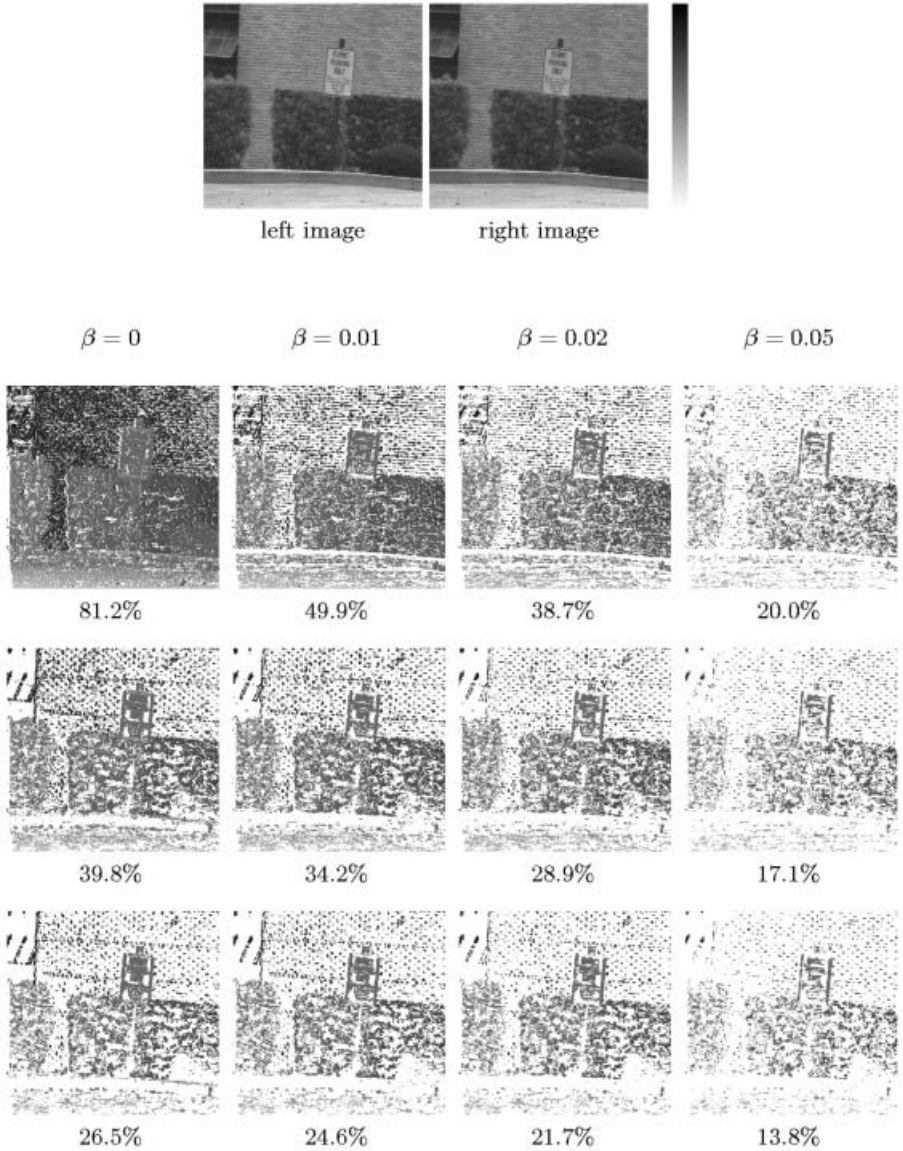


**Fig. 3.** Matching results on the stripes scene. Left image of high and low contrasts (1st column), disparity maps (2nd column) and segmentations (3rd column) by confidently stable matching with  $\alpha = 10$ ,  $\beta = 0.02$  (segmentation colors: green—matched; blue—half-occluded; red—ambiguous, many matches possible; yellow—ambiguous, matched/occluded uncertainty), disparity map by the MAP algorithm is shown for comparison (last column).

a thin pole just in front of a tall hedge. The leafed shrubs cannot be considered as a surface at the scale of image resolution, so no continuity prior can be used. Five shrubs can be distinguished, they differ by their distance from the camera. The roadside is imaged at high inclination angle relative to the camera pair, so in addition to the lack of texture (Class 1 ambiguity) it also exhibits large relative affine distortion between the images. The rim of the curb is parallel to epipolar lines in the left half of the image and it is thus difficult to match (Class 2 ambiguity). The background wall images as a very regular texture (Class 2 ambiguity) and it is thus difficult to match without a continuity prior. The dark space in the open window does not contain any stereoscopic information. Besides having no natural texture, the window glass partly reflects the featureless cloudy sky (both Class 1 ambiguity). The images do not exactly correspond to the same scene because of time difference breeze artefacts in the image capture.

Disparity maps shown in Fig. 4 are given for varying  $\alpha$  (constant in rows) and  $\beta$  (constant in columns). In the first row ( $\alpha = 0$ ) we can see that  $\beta$  itself cannot resolve ambiguity due to low SNR (Class 1): the noisiness in the shadowed window area remains until large values of  $\beta$  (2.5% of signature similarity range). The structural ambiguity (Class 2) is detected quite early: the number of mismatches on the brick wall and on the rim of the curb decreases fast with increasing  $\beta$ .

In the first column ( $\beta = 0$ ) we can see that  $\alpha$  itself does not detect structural ambiguity very well: large proportion of mismatches remain on the wall and the curb rim. The Class 1 ambiguity is detected early, on the other hand.



**Fig. 4.** The Shrub image pair (the wide-most stereo baseline of the set). Matching results for varying  $\alpha$  (per rows:  $\alpha = 0$ ,  $\alpha = 10$ ,  $\alpha = 20$ ) and  $\beta$  (columns). Disparity color coding is shown beside the images (light grey: high disparity, black: low disparity, white: holes). Matching window  $5 \times 5$ , disparity search range (10, 30), image size  $480 \times 512$ , mean CPU time 8.2 sec (P7, 1GHz), percentage of matched pixels given under maps. See the PDF version with color-coded maps where matching errors are more apparent.

We conclude  $\alpha$  is related to the ability to detect Class 1 ambiguity and  $\beta$  is related to the ability to detect Class 2 ambiguity. Both have to be non-zero. Increasing both of them decreases disparity map density but also eliminates errors. Our short experience with the algorithm shows that  $\alpha = 10$  and  $\beta = 0.02$  is suitable for a broad range of natural scenes. Under high structural ambiguity, increasing  $\beta$  up to 0.04 suffices.

Note that a few points of high disparity remain in the bushes for even very wide confidence intervals. These are due to the wind artefacts mentioned above.

Another example of an outdoor scene is shown Fig. 5. The building is the Branitz Castle in Brandenburg, Germany (built 1772). The image is a section of a 360° panorama captured by a non-perspective camera with two vertically aligned line sensors placed above each other and rotating around a common vertical axis [31]. Epipolar lines are vertical in these images.



**Fig. 5.** Left image and disparity map of Branitz Castle. Disparity coding is the same as in Fig. 4. Matching window  $7 \times 7$  pixels,  $\alpha = 10$ ,  $\beta = 0.02$ , disparity search range (0, 90), image size  $2037 \times 4421$ , 28% matched.

This example was selected because of great scene depth and because it exhibits various combinations of Class 1 and Class 2 ambiguity.

The disparity map is shown in Fig. 5. A few significant isolated errors appear on the pilasters and near the bottom image edge. Even small and thin objects are matched: the flower pots, the lions, and the lamp posts, the tree trunk supported by the three pales, the wire loop on the flagpole or the tips of the lightning rods on the roof corners and the chimneys. The individual branches of the foreground tree are well distinguished. Ambiguous regions are suppressed, most notably on the façade and the staircase. Note the tiled roof is matched quite densely; the stripes of false negatives are sampling artefacts (due to spatial frequencies close to Nyquist limit). Note the disparity map is especially dense on the lawn, the shrubbery alongside the building, and the foliage in the background, i.e. in regions of irregular texture.

We conclude the combination of Class 2 and Class 1 ambiguity is the worst problem. But the Class 2 ambiguity can be reduced by using more views. With an independent view the disparity map density would increase significantly.

To partially compare confidently stable matching with other stereo matching algorithms we run Alg. 1 on the Scharstein & Szeliski’s test set [32]. The error was evaluated only on the set of matches found by the algorithm. The results shown in Fig. 6 were computed for  $\alpha = 10$ ,  $\beta = 0.02$ ,  $5 \times 5$  matching widow using correlation (2) and (3). The average running time (CPU 1GHz P7) was 3.1 sec. We can see a straightforward stereo matching implementation based on the stability principle alone can already be ranked as the fourth best (Feb 2002).



dataset	map density	error statistics		
		$B_{\bar{O}}$	$B_{\bar{T}}$	$B_{\mathcal{D}}$
tsukuba	45%	1.4%	0.22%	7.8%
sawtooth	52%	1.6%	0.03%	14.0%
venus	40%	0.8%	0.04%	10.0%
map	74%	0.3%	0.20%	3.6%

**Fig. 6.** Results on the Middlebury data set.  $B_{\bar{O}}$  is the percentage of mismatches in non-occluded region,  $B_{\bar{T}}$  is the number of mismatches in textureless areas, and  $B_{\mathcal{D}}$  is the number of errors at disparity discontinuities [32].

## 5 Conclusions

We have defined confidently stable matching that is able to (1) find the largest unambiguous component of stereo matching and (2) to identify the part of solution where regularization is necessary to resolve ambiguities.

We have seen on three natural scenes and on three artificial scenes that the unique component of stable matching is large (10% to 80% of image density) even if confidence interval for match similarity measure is wide (2.5% of its range). Moreover, the experiment shows that confidently stable matching has very low error rate (0.3% to 1.4% of all matches) and that even very small objects are correctly matched.

The advantage of stability principle is that it gives a great freedom in defining what is a set of good matches. Most importantly, since stability is a property of partial orderings, it is easy to extend the results to multi-criterial matching, which opens a possibility to influence the low-level vision process by some high-level scene understanding process by means of meta-rules defining the relative importance of the individual criteria. This is a topic for our future research.

**Acknowledgements.** The author was supported by the Grant Agency of the Czech Republic under project GACR 102/01/1371 and by the Czech Ministry of Education under project MSM 210000012. Part of this work was conducted while the author was visiting the CITR group at the University of Auckland. The support of CITR is gratefully acknowledged. The Branitz Castle images are courtesy of DLR Berlin, the Shrub images are courtesy of Carnegie Mellon University.

## References

1. Julesz, B.: Towards the automation of binocular depth perception (AUTOMAP-1). In: IFIPS Congress, Munich (1962)
2. Kutulakos, K.N., Seitz, S.M.: A theory of shape by shape carving. *IJCV* **38** (2000) 199–218
3. Baker, S., Sim, T., Kanade, T.: A characterization of inherent stereo ambiguities. In: *Proc ICCV*. (2001) 428–435
4. Schlesinger, M.I.: Ambiguity in stereopsis. Personal communication (1998)
5. Šára, R.: Failure analysis of stable matchings. Research report (In preparation)
6. Stewart, C.V., Dyer, C.R.: The trinocular general support algorithm: A three-camera stereo algorithm for overcoming binocular matching errors. In: *Proc ICCV*. (1988) 134–138
7. Satoh, K., Ohta, Y.: Occlusion detectable stereo using a camera matrix. In: *Proc ACCV*. (1995) 331–335
8. Marr, D.: A note on the computation of binocular disparity in a symbolic, low-level visual processor. A.I. Memo 327, AI Lab, MIT (1974)
9. Pollard, S.B., Mayhew, J.E.W., Frisby, J.P.: PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception* **14** (1985) 449–470
10. Zitnick, C.L., Kanade, T.: A cooperative algorithm for stereo matching and occlusion detection. *IEEE Trans PAMI* **22** (2000) 675–684

11. Belhumeur, P.N.: A Bayesian approach to binocular stereopsis. *IJCV* **19** (1996) 237–260
12. Bobick, A.F., Intille, S.S.: Large occlusion stereo. *IJCV* **33** (1999) 181–200
13. Robert, L., Deriche, R.: Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In: Proc ICIP. (1992) 123–127
14. Barnard, S.T.: Stochastic stereo matching over scale. *IJCV* **3** (1989) 17–32
15. Scharstein, D., Szeliski, R.: Stereo matching with nonlinear diffusion. *IJCV* **28** (1998) 155–174
16. Boykov, Y., Veksler, O., Zabih, R.: Disparity component matching for visual correspondence. In: Proc Conf CVPR. (1997) 470–475
17. Ishikawa, H., Geiger, D.: Occlusions, discontinuities, and epipolar lines in stereo. In: ECCV. (1998) 232–248
18. Roy, S., Cox, I.J.: A maximum-flow formulation of the  $n$ -camera stereo correspondence problem. In: Proc ICCV. (1998) 492–499
19. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions using graph cuts. In: Proc ICCV. (2001) 508–515
20. March, R.: Computation of stereo disparity using regularization. *Pattern Recognition Letters* **8** (1988) 181–187
21. March, R.: A regularization model for stereo vision with controlled continuity. *Pattern Recognition Letters* **10** (1989) 259–263
22. Jung, D.Y., Oh, J.H., Lee, S.C., Lee, C.H., Nam, K.G.: Stereo matching by discontinuity-preserving regularization. *J of Elect Eng and Inf Sci* **4** (1999) 452–8
23. Manduchi, R.; Tomasi, C.: Distinctiveness maps for image matching. In: Proc ICIAP. (1999) 26–31
24. Šára, R.: Sigma-delta stable matching for computational stereopsis. Research Report CTU–CMP–2001–25, Center for Machine Perception, Czech Technical University (2001)  
[<ftp://cmp.felk.cvut.cz/pub/cmp/articles/sara/Sara-TR-2001-25.pdf>].
25. Krol, J.D., van der Grind, W.A.: Rehabilitation of a classical notion of Panum’s fusional area. *Perception* **11** (1982) 615–619
26. Yuille, A.L., Poggio, T.: A generalized ordering constraint for stereo correspondence. A.I. Memo 777, AI Lab, MIT (1984)
27. Šára, R.: A fast algorithm for confidently stable matching. Research Report CTU–CMP–2002–03, Center for Machine Perception, Czech Technical University (2002)  
[<ftp://cmp.felk.cvut.cz/pub/cmp/articles/sara/Sara-TR-2002-03.pdf>].
28. Moravec, H.P.: Towards automatic visual obstacle avoidance. In: Proc IJCAI. (1977) 584
29. Mandelbaum, R., Kamberova, G., Mintz, M.: Stereo depth estimation: a confidence interval approach. In: Proc ICCV. (1998) 503–509
30. Cox, I.J., Hingorani, S., Maggs, B.M., Rao, S.B.: Stereo without disparity gradient smoothing: a Bayesian sensor fusion solution. In: Proc BMVC. (1992) 337–346
31. Scheibe, K., Korsitzky, H., Reulke, R., Scheele, M., Solbrig, M.: EYESCAN—a high resolution digital panoramic camera. In: Int Wkshp Robot Vision, Auckland (2001) 77–83
32. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Technical Report MSR–TR–2001–81, Microsoft Research, Redmont, WA (2001) To appear in *IJCV*.