CENTER FOR
MACHINE PERCEPTION

CZECH TECHNICAL
UNIVERSITY

RESEARCH REPORT

ISSN 1213-2365

Sigma-Delta Stable Matching for Computational Stereopsis

(Version 1.5)

Radim Šára

sara@cmp.felk.cvut.cz

CTU–CMP–2001–25

September 3, 2001

Available at
ftp://cmp.felk.cvut.cz/pub/cmp/articles/sara/Sara-TR-2001-25.pdf

This research has been supported by Grant Agency of the Czech Republic under grant GACR 102/01/1371 and by Czech Ministry of Education under grant MSM 212300013.

Research Reports of CMP, Czech Technical University in Prague, No. 25, 2001

Published by

Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering, Czech Technical University
Technická 2, 166 27 Prague 6, Czech Republic
fax +420 2 2435 7385, phone +420 2 2435 7637, www: http://cmp.felk.cvut.cz

Abstract

In this paper we explore stability principle applied to the problem of stereoscopic matching. Stability is based on competition among candidate matches. It is not an optimality condition, it is a matching property, much like ordering or uniqueness. Unlike the latter two it guarantees the existence of a unique solution to a matching problem. Very roughly speaking, stable matching is one in which each selected match dominates its potential competitors in some property. Great versatility is possible since the choice of the property is only very weakly constrained. The simplest property of matches, explored in this paper, is their similarity measure computed over small rectangular image neighborhoods as in standard area-based matching.

We define stability, give the existence and uniqueness theorem, and suggest a simple and fast algorithm. We then show how stable matching copes with wide disparity search range, half-occluded regions, very thin objects, and featureless areas. We will show how stability property extends naturally to the case when confidence intervals instead of single-valued match similarities are given. This allows for guaranteed high accuracy in stereoscopic matching. In a ground-truth experiment we compare various matchings in terms of false positives, false negatives, mismatches, and bias towards large objects.

Contents

1 Introduction

Stereo matching is one of the inverse problems long studied in computer vision [9]. Because of its ill-posedness [1] much work has been devoted to posing the problem such that a prior continuity model could be incorporated while keeping the computational complexity of the matching algorithm low. The early solutions based on the heuristic of *cohesiveness* [14] culminated in the classical algorithm by Pollard, Mayhew and Frisby [16]. Although some interest remained [23], 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 Belhumeur [2] (assuming continuity along epipolar lines) and by Robert and Deriche [17] (assuming isotropic continuity prior). Recent disparity component matching [4] and network flow [8, 18, 10] formulations of the matching task also include isotropic prior model but their computational complexity is lower. The most promising recent results are due to Kolmogorov and Zabih [10]. They formulate the matching problem as constrained minimization of Potts energy functional including a data conformance term, a continuity term and half-occlusion penalty term. The optimum is searched for under uniqueness constraint. The detection of half-occluded regions is based on the relative ratio of data conformance vs. occlusion penalty and is not supported by any geometric constraint. The authors give two approximate optimum search algorithms of which their α-expansion procedure works better.

The difficulty in stereo matching using any form of prior continuity models is that discontinuities and half-occlusions must be accounted for properly. Otherwise heavy artifacts may result from the simultaneous occurrence of depth discontinuity *and* low signal-to-noise ratio. This happens when 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.

Here we will be interested in algorithms that are able to *decide* whether evidence in data is sufficient to establish stereo correspondences. The decision must be based on the data itself, since global models (thresholds) are known to work poorly.

The ability to reject unreliable and/or ambiguous matches is extremely important when information from multiple views is to be integrated into a globally consistent geometric model. The interpolation, on the other hand, is a higher-level problem and can be solved within the geometric model reconstruction procedure *provided that the matches found are reliable enough.*

Therefore, the question we would like to open can be stated as follows: *How information available in the input images can be used efficiently to find a matching without a continuity prior?* It is clear that in such a case one must rely on just the *unambiguous* evidence available in input data. It is also clear that one cannot aim at obtaining dense matchings: if images contain no information about correspondences (images of the clear sky) we require the matching algorithm to assign *no matches.*

Let us develop a bit of insight: 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 conditions the (actual) matches dominate all other candidate matches in their signature similarity measure.[1] It is then easy to find such dominant pairs. We call this *dominant matching.* A special case of dominant matching is the popular 'symmetric maximum method.' In the noise-free case dominant matching is complete[2] and leaves only half-occluded regions unmatched. Dominant matching will be defined precisely in Section 3.

If signatures are *corrupted by random noise*, our correlation value becomes a random variable. Some correct correspondences receive lower correlation and some other candidate correspondences receive higher correlation. Since there are usually many competitors to a given correct match, the probability of a false match becoming dominant is

[1] Pixel signature similarity will be called correlation from now on.
[2] Complete matching is 'as dense as possible', see later for precise definition.

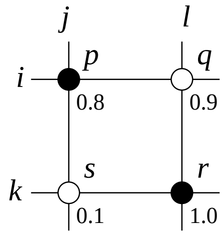


Figure 1: Let $M = \{p, r\}$ be a matching (full circles) on a 2×2 matching problem. Match $p = (i, j)$ of correlation 0.8 has a competitor $q = (i, l)$ of correlation 0.9. The q has a competitor $r = (k, l)$ of correlation 1.0. The pair p is thus stable in M (its competitor has a competitor from M).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

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. The question addressed in this paper is whether we may weaken the dominance constraint to improve the density of the resulting matching while keeping the probability of mismatch low.

The answer is positive. We weaken dominance to *stability*. Roughly speaking, matching is stable if each match p included in the matching has no other competitors q besides those that have a competitor r which *is* a member of the matching (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 or ordering constraints, for instance) *and* the correlation of q is greater than that of p (precise definition will be given in Section 3). Obviously, the q would be a better candidate to be matched on account of its correlation value alone.

Like the well-known ordering constraint, stability only requires certain condition to hold. Remarkably enough, stable matching is unique and, unlike the dominant matching, it is also *complete* (under mild assumptions stated later), even under noise.

The disadvantage of the stability introduced above is its susceptibility to noise: the ordering of potential matches, being based on their correlation value, can get easily disturbed. We therefore generalize stability to $\sigma\delta$-stability which allows for some margin when comparing correlations. The $\sigma\delta$-stable matchings are no longer complete but they are remarkably error free. Both dominant and stable matchings are special cases of $\sigma\delta$-stable matchings. Section 4 deals with this most general form of stability and it is where the main contribution of this paper lies. In Section 5 we use it to formalize the problem of stereo matching when *confidence intervals* as opposed to single correlation values are given for all potential matches.

The Stable Matching Problem studied here is similar to the Stable Marriage Problem [7] which has been studied intensively since Gale and Shapley published their classical paper [6]. The Stable Marriage Problem formulation is more general in certain respects (it allows for asymmetric rankings of preferences) and less general in the other (rankings are less general than partial orderings).

In this paper we define stable matching for the special case when all potential matches can be globally ranked (with ties). This is often the case in area-based stereo using similarity measure like the sum of squared differences or normalized cross-correlation. The more general possibility of defining stability on partial orderings is only hinted here as it is still a topic for our ongoing research.

We discuss only binocular matching problem here as it is the most difficult case. We will not deal with polynocular stereo in this paper, since the generalization is simple, if not straightforward.

2 Basic Concepts

We first briefly review the principal concepts upon which the notion of stability is build.

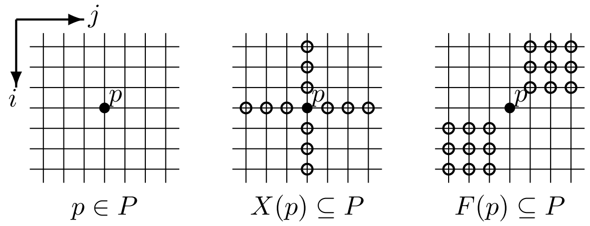


Figure 2: A pair p in matching table P (*left*), the X-zone $X(p)$ (*center, empty circles*), and the F-zone $F(p)$ (*right*).

Matching table Let $I, J = \{1, 2, \ldots, N\}$ 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 $p = (i, j) \in P$ will be called a (candidate) *pair*. Let c be correlation measure assigning each pair p a real number $c(p)$. We will visualize P as a *matching table* as shown in Fig. 2 in which each crossing (circled or not) may represent a pair from P. Note that the $I \times J$ may or may not be fully populated by P.

The X-zone and the FX-zone The *X-zone*[3] $X(p) \subseteq P$ 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*[4] $F(p) \subseteq P$ 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, $FX(p) = X(p) \cup F(p)$.

Matching 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' [14]).

The *cardinality* of matching M is denoted as $|M|$. A maximum cardinality matching has the greatest number of pairs possible.

Monotonic matching 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' [22]).

Inhibition zone Let us, for a moment, consider the general subsets M of P (not necessarily matchings). Each 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. Two pairs that are not discordant will be called concordant.

For instance, X-zone is the inhibition zone for matchings and F-zone is the inhibition zone for monotonic subsets of P. Therefore, the subset M is a matching iff $X(p) \subseteq Z(p)$ for each $p \in M$ and is a monotonic matching iff $FX(p) \subseteq Z(p)$ for each $p \in M$.

We can see that both uniqueness and ordering constraints have a very similar representation. Since the early work on stereo matching little has been done to incorporate the zones in a natural way into the matching problem task formulation.

Inhibition zone shape 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.

[3]The name of the X-zone was chosen because of its shape.

[4]The F-zone stands for *forbidden zone* [11].

Inhibition zone depth The zone depth D is defined by the largest relative disparity of the pairs within, as follows:

$$D(Z(p)) = \max_{q \in Z(p)} \max(|i-k|, |j-l|),$$

where $q = (k, l)$ and $p = (i, j)$. See Fig. 2, where the F-zone shown is of depth $D = 3$.

Matching problem Let P be matching table (sparsely or fully populated), $c\colon P \to \mathbb{R}$ be a bounded correlation function defined on the table and $Z(p)$ be inhibition zones for each $p \in P$ such that they generate a symmetric discordance relation (i.e., all of them are symmetric and of the same shape and depth[5]). The triple (P, c, Z) will be called a matching problem.

3 Stable Matching

In this section we define stable matching. Stability is a property of subsets of matching table P and is defined on (partial) ordering relation on pairs from P. The ordering is generated by the competition relation which we define here as follows:

Definition 1 (Competitor). *We say that pair $q \in P$ is a competitor to pair $p \in P$ iff $q \in Z(p)$ and $c(q) \geq c(p)$. We say it is a strict competitor if, in addition, $c(q) > c(p)$.*

Stability is now defined as follows:

Definition 2 (Stability). *Let (P, c, Z) be a given matching problem and let $M \subseteq P$. We say a pair $p \in P$ is stable with respect to M iff for each $q \in P$ which is a competitor to p there is a pair $r \in M$ which is a strict competitor to q. The subset M is stable if every $p \in M$ is stable with respect to M.*

We will refer to q as the *direct* competitor to p and to r as the *indirect* competitor to p. To stress the fact that a particular inhibition zone Z is generating a stable subset of P, we say such subset is Z-stable.

In stereo matching there is a popular simple method of matching two sets of pixels (or interest/feature points in wide-baseline stereo), it is usually called symmetric maximum matching: pixel $i \in I$ is matched to pixel $j \in J$ if the correlation $c(i, j)$ is greater than the correlation $c(i, k)$ for any other $k \in J$ *and*, symmetrically, the correlation $c(i, j)$ is greater than the correlation $c(l, j)$ for any other $l \in I$. Here we call such matching X-dominant. Using our terminology, dominant matching for general inhibition zones is defined as follows:

Definition 3 (Dominance). *Let (P, c, Z) be a given matching problem. A pair $p \in P$ is called Z-dominant if it has no competitors in $Z(p)$. A subset $M \subseteq P$ is Z-dominant if every $p \in M$ is Z-dominant.*

As before, such subset M is a matching iff $X \subseteq Z$ and is monotonic iff $F \subseteq Z$. Moreover, the following holds: Let $D \subseteq P$ be Z-dominant subset on (P, c, Z) and let $S \subseteq P$ be Z-stable subset on (P, c, Z). Then $D \subseteq S$. A proof for the case $Z = FX$ is given in [19]. Both dominant matching and stable matching are special cases of $\sigma\delta$-stable matching to be discussed in Section 4.

Dominant matchings are known to be rather sparse. This was the reason that led us to the formulation of a condition which is weaker than dominance and which guarantees greater matching density. To formalize the somewhat vague notion of matching density we introduce matching completeness as follows:

Definition 4 (Completeness). *A subset $M \subseteq P$ is Z-complete iff for each pair $p \notin M$ it holds that the intersection $Z(p) \cap M$ is non-empty.*

[5]Zones close to the border of the matching table may extend past the boundary of the table. We simply assume there are no candidate pairs outside of the table.

The Z-stable matching is Z-complete if there are no two discordant pairs of equal cost (which is rare). The Z-dominant matching is generally (quite often) incomplete. Complete matchings give denser disparity maps than matchings that are not complete. Experimental comparison of disparity map density for several stable matching algorithms will be given in Section 6.

3.1 Quasi-Stable Matching Algorithm

We first give a simple matching algorithm and prove its correctness. This algorithm ignores discordant ties (equal-cost discordant pairs), which is the reason we call it quasi-stable. Quasi-stable matchings have important properties which will be discussed shortly. Namely, it will be shown that Z-stable matching is a subset of quasi-Z-stable matching and can be found by post-processing using Alg. 2. This combination has the best time complexity of all algorithms for stable matching we have explored so far.

Algorithm 1 (Quasi-Stable Matching).

Input: *Matching problem* (P, c, Z).

Output: *Stable matching* M.

Procedure:

1. *Initialize* $M := \emptyset$.
2. *If* P *is empty, terminate.*
3. $M := M \cup \{p\}$, *where* p *is the largest-cost pair in* P. *If there are multiple such pairs, select any of them.*
4. *Remove* $Z(p) \cup \{p\}$ *from* P *and go to 2.*

The proof of Alg. 1 correctness will be stated by Theorem 1, based on the following three lemmas:

Lemma 1. *Let* Z *be a symmetric inhibition zone. Any run of Alg. 1 on* (P, c, Z) *returns a complete matching.*

Proof. Suppose the set M returned by the algorithm is not complete. Then there is a pair $p \in P \setminus M$ such that $Z(p) \cap M = \emptyset$.[6] By symmetry of inhibition zone it must hold for each pair $q \in M$ that $p \notin Z(q)$. But this is a contradiction, since, in this case the p would have been assigned to M in Step 3 of Alg. 1. □

The following two lemmas are general characterizations of stable complete matchings irrespective of the existence of a matching algorithm:

Lemma 2. *Let* (P, c, Z) *be a matching problem instance with no discordant ties. Any stable complete matching* $\bar{M}$ *on* (P, c, Z) *contains the highest-correlation pair in* P.

Proof. Suppose there is such matching $\bar{M} \subseteq P$ and the highest-correlation pair $p \in P$ is not a member of $\bar{M}$. From the requirement of completeness, it follows there is a pair $q \in Z(p) \cap \bar{M}$ such that $c(q) < c(p)$. But if q is to be stable there must be a pair $r \in Z(p) \cap \bar{M}$ such that $c(r) > c(p)$, which is a contradiction. □

The following lemma suggests how a stable complete matching can be decomposed into a series of nested subproblems:

Lemma 3. *Let* (P_1, c, Z) *be a matching problem without discordant ties. Let a matching subproblem* (P_2, c, Z) *be constructed such that* $P_2 = P_1 \setminus (Z(p_1) \cup \{p_1\})$, *where* p_1 *is the highest-rank pair in* P_1. *Then the following holds:*

1. *If* $\bar{M}_1$ *is a stable complete matching for* (P_1, c, Z) *then* $\bar{M}_2 = \bar{M}_1 \setminus \{p_1\}$ *is a stable complete matching for* (P_2, c, Z).

[6] The $\setminus$ is a set-theoretic difference.

2. *If $\bar{N}_2$ is a stable complete matching for (P_2, c, Z) then $\bar{N}_1 = \bar{N}_2 \cup \{p_1\}$ is a stable complete matching for (P_1, c, Z).*

Proof. The first statement holds, since after deleting $Z(p_1)$ from P_1, both 'stable' and 'complete' properties of $\bar{M}_2$ are inherited from $\bar{M}_1$. The second statement holds, since correlation values in $Z(p_1)$ are always smaller than $c(p_1)$, thus the 'stable' property holds, and the completeness is obvious. □

As a corollary of Lemma 3 we have that for a given matching problem (P, c, Z) without discordant ties there is always a *unique* stable complete matching.

Theorem 1. *For any matching problem instance without discordant pairs, Alg. 1 terminates, and returns a unique stable complete matching.*

Proof. We construct a series of strictly nested problems as in Lemma 3. After n steps, $n = \min(|I|, |J|)$, we end up with a trivial matching problem (P_n, r), for which stable complete matching is $\bar{M}_n$. By Lemma 3, the $\bar{M}_1$ is constructed as $\bar{M}_n \cup \{p_n\} \cup \{p_{n-1}\} \cup \cdots \cup \{p_1\}$ and is a stable complete matching. Since the process always succeeds in solving the problem P_n, there is always a stable matching. Because of the nesting strictness, there is at most one such matching. □

We will now release the assumption of no discordant ties. It is easy to see that the set M returned by any run of Alg. 1 has the following property:

Property 1. *For each $p \in M$ and $q \in Z(p)$ such that $c(q) \geq c(p)$ it holds that there is a pair $r \in Z(q) \cap M$ such that $c(r) \geq c(q)$.*

This is the reason we call the result of Alg. 1 quasi-stable matching. There may be more quasi-stable matchings for a given matching problem (P, c, Z).

We will not discuss stable matching algorithm for the case when ties are allowed because it is a special case of algorithm for $\sigma\delta$-stable matching which is defined and discussed in the next section.

Time Complexity The time complexity of Alg. 1 is $O(|P| \cdot |M|)$, where $|P|$ is the number of initial pairs in the matching table and $|M|$ is the cardinality of the algorithm's output. If P is an $n \times n$ densely populated table the algorithm complexity is $O(n^3)$. It may be reduced to $O(n^2 \log n)$ when pairs waiting to be considered by the algorithm are represented as a heap of column (or row) heaps and the pairs rejected in Step 4 are represented by binary search trees in each column (row) of the table.

4 $\sigma\delta$-Stable Matching

The fact that the largest-cost pair is a member of stable matching (unless it is discordant with another pair of the same cost) may be viewed as a property which is not always desirable. Correlation computed from image data is a random variable: correlation measure may take on any value within its range for a pair of images of a real surface point neighborhood as well as for a pair of non-corresponding point neighborhoods. In practice, this happens quite often and one faces the situation when a pair which obviously does not correspond to any real surface patch has a rather high correlation value. The influence of such pairs upon the matching could be reduced if we allowed for some margin when comparing two correlations in the stability definition. We explore this possibility here.

Definition 5 (ϵ-Competitor). *Pair $q \in P$ is an ϵ-competitor to $p \in P$ iff $q \in Z(p)$ and $c(q) \geq c(p) - \epsilon$ and is a strict ϵ-competitor if, in addition, $c(q) > c(p) - \epsilon$.*

Definition 6 ($\sigma\delta$-Stability). *Let σ and δ be two constants such that $-\delta \geq \sigma \geq 0$. Pair $p \in P$ is $\sigma\delta$-stable with respect to $M \subseteq P$ iff for each $q \in P$ which is a σ-competitor to p there is an $r \in M$, $r \neq p$, which is a strict δ-competitor to q. Matching M is $\sigma\delta$-stable iff each $p \in M$ is $\sigma\delta$-stable with respect to M.*

We say the above triple (p, q, r) verifies stability. In the triple, the pairs p, r will be called the tested pairs and the q will be called the query pair.

Comments

1. Note that when $\epsilon = \infty$ then any pair from inhibition zone $Z(p)$ is an ϵ-competitor to p, when $\epsilon = -\infty$ then no pair is a competitor to p. Various combinations of extreme values for σ and δ lead to different types of stable matching as can be seen from the following table:

σ	δ	matching
0	0	stable
0	$-\infty$	dominant
∞	$-\infty$	empty

We can see that the matching is allowed to vary from empty to full cardinality when σ and δ vary within their range.

2. If we did not restrict δ to $\delta \leq 0$, there might be more solutions to a given $\sigma\delta$-stable matching problem.

3. If we did not impose the restriction $-\delta \geq \sigma$ and we only required $\delta \leq 0$ we would need to require the indirect competitor to a pair p to be different from p. Otherwise computational problems would arise (the matching problem would become NP complete).

4. Every pair $p \in P$ can be associated with different σ and δ, which can be related to the confidence interval for $c(p)$, for instance. More on this possibility in Section 5.

4.1 $\sigma\delta$-Stability Properties

We are now ready to prove the following lemma:

Lemma 4. *Let $\sigma \geq 0$ and $\delta \leq 0$ and let Z be a symmetric inhibition zone. Then the $\sigma\delta$-stable matching on (P, c, Z) is a subset of any quasi-stable matching on (P, c, Z).*

Proof. Let M be some quasi-stable matching on (P, c, Z) and let S be $\sigma\delta$-stable matching on (P, c, Z). We prove the lemma by contradiction. Let $s_0 \in S$ and $s_0 \notin M$. From the fact that $s_0 \notin M$ and from completeness of M it follows that there is a pair $t_0 \in Z(s_0)$ such that $t_0 \in M$. There are three mutually exclusive cases:

1. $c(t_0) > c(s_0)$. Then, because t_0 is a σ-competitor to s_0 (since $\sigma \geq 0$), there must be $s_1 \in Z(t_0) \cap S$ such that $c(s_1) > c(t_0) - \delta \geq c(t_0)$ to preserve the $\sigma\delta$-stability of s_0. Since $t_0 \in M$ it holds that $s_1 \notin M$ since $s_1 \in Z(t_0)$. In order for t_0 to be in M there must exist $t_1 \in Z(s_1) \cap M$ such that $c(t_1) \geq c(s_1)$ (by Property 1 above). Since $\delta \leq 0$ the argument continues along a strictly non-decreasing path until
   (a) There is no such t_i. Then $t_{i-1} \in M$ does not satisfy Property 1 which is a contradiction.
   (b) There is no such s_{i+1}. Then s_i is not $\sigma\delta$-stable but $s_i \in S$, which is a contradiction.

2. $c(t_0) < c(s_0)$. Then, since $t_0 \in M$, there must be $t_1 \in Z(s_0) \cap M$ such that $c(t_1) \geq c(s_0)$. Since $\sigma \geq 0$ and s_0 is $\sigma\delta$-stable there must be $s_1 \in Z(t_1) \cap S$ such that $c(s_1) > c(t_1) - \delta \geq c(t_1)$. The argument continues the same way as in the previous case and also leads to contradiction.

3. $c(t_0) = c(s_0)$. Then, since s_0 is $\sigma\delta$-stable, $\sigma \geq 0$, and $\delta \leq 0$ there must be $s'_0 \in Z(t_0) \cap S$ such that $c(s'_0) > c(s_0)$ and we thus have Case 2.

□

4.2 $\sigma\delta$-Stable Matching Algorithm

Before we introduce the $\sigma\delta$-stable matching algorithm we first slightly extend the notion of $\sigma\delta$-stability. Let $\delta : P \times P \to \mathbb{R}$ and $\sigma : P \times P \to \mathbb{R}$ be two scalar bounded functions on couples of pairs. The stability definition will now read as follows

Definition 7 (Generalized $\sigma\delta$-stability). *Matching S is $\sigma\delta$-stable if for each $p \in S$ and each $q \in Z(p)$ the following holds:*

$$\text{If } c(q) \geq c(p) - \sigma(p,q) \text{ then there must be } r \in S \text{ such that } c(r) > c(q) - \delta(q,r). \quad (1)$$

The functions σ, δ are not symmetric, in general. We will distinguish two important cases

1. When $\sigma(p,q) = \sigma(p)$ and $\delta(q,r) = \delta(r)$ we say the σ, δ functions are associated with tested pairs in each triple (p,q,r) verifying stability.

2. When $\sigma(p,q) = \sigma(q)$ and $\delta(q,r) = \delta(q)$ we say they are associated with the query pairs.

Converting Pairs We are now interested in converting quasi-stable matching M to $\sigma\delta$-stable matching by removing some pairs from M.

Pair q participates in converting quasi-stable matching M to $\sigma\delta$-stable matching S by removing pair $s \in M \cap Z(q)$ iff $c(q) \geq c(s) - \sigma$ and there is no pair $p \in S \cap Z(q)$ such that $c(p) > c(q) - \delta$. Note it follows that $q \notin M$ and $p \neq s$.

Let M be a quasi-stable matching on (P, c, Z). Let $q \in P$. We introduce two sets $T(q \mid Q)$ and $D(q \mid R)$ given sets $Q, R \subseteq M$ as follows:

$$T(q \mid Q) = \{t \mid t \in Z(q) \cap Q,\ c(t) > c(q) - \delta(q,t)\}\,, \quad (2)$$

$$D(q \mid R) = \{d \mid d \in Z(q) \cap R,\ c(d) \leq c(q) + \sigma(d,q)\}\,. \quad (3)$$

Note that if $q \in M$ then $T(q \mid M) = D(q \mid M) = \emptyset$. We will now rephrase the converting pair condition as follows:

Observation 1. *Let S be $\sigma\delta$-stable matching for matching problem (P, c, Z). Let $C \subseteq P$ be the set of pairs that convert quasi-stable matching M to S. Then, for each $q \in C$ it holds that $T(q \mid S) = \emptyset$ and $D(q \mid M) \notin S$.*

Before we come to the algorithm let us impose a condition on σ, δ: Let $q, t, q' \in P$, $t \notin \{q, q'\}$. We require that

$$\sigma(t, q') + \delta(q, t) \leq 0 \text{ when } c(q) \geq c(q'). \quad (4)$$

It is worth noting that this condition includes several special cases:

1. When $\sigma = 0$ and $\delta(q,t) \leq 0$. The extreme cases are (ordinary) Z-stable matching when $\delta = 0$ and Z-dominant matching when $\delta = -\infty$.

2. When $\sigma = const$ and $\delta = const$ and $\sigma \leq -\delta$.

3. When $\sigma(t, q') = -\delta(q, t)$, for all $q, q' \in Z(t)$, i.e., when σ and δ are associated with tested pairs in each triple that verifies $\sigma\delta$-stability and $\sigma(t, q') = \sigma(t)$, $\delta(t) = \delta(q, t)$. This case is important, since the confidently stable matching problem (to be defined in Section 5) can be reduced to this case.

Under condition (4) the $\sigma\delta$-stable matching algorithm works as follows:

Algorithm 2 ($\sigma\delta$-Stable Matching).

Input:
- *matching problem* (P, c, Z),
- *functions* $\sigma, \delta \colon P \times P \to \mathbb{R}$ *satisfying* (4).

Output: *$\sigma\delta$-stable matching S*

Procedure:

1. *Run Alg. 1 to obtain any quasi-stable matching $M \subseteq P$.*
2. *Set $i := 0$ and initialize $S_i := M$ and $L_i := P \setminus M$.*
3. *Sort L_i in descending order with respect to c.*
4. *Until L_i is empty repeat 5–7.*
5. *$q_i := \mathrm{head}(L_i)$, $L_{i+1} := \mathrm{tail}(L_i)$.*
6. *if $T(q_i \mid S_i) = \emptyset$ then $S_{i+1} := S_i \setminus D(q_i \mid M)$ else $S_{i+1} := S_i$.*
7. *$i := i + 1$.*

Comments

1. It is possible to assign $L_i := P$ in Step 2 instead of $P \setminus M$ since $D(p \mid M) = \emptyset$ for all $p \in M$.
2. L may be reduced by eliminating all pairs for which $D(p \mid M) = \emptyset$, where M is quasi-stable matching. It can be done in $O(|P|)$ time if $Z \in \{X, F, FX\}$.
3. The test in Step 6 can be speeded up as follows. For each pair $q \in P$ we determine the largest-cost pair $p \in Z(q) \cap M$. If p remains in S_i at the time of testing q in Step 6, then the test is equivalent to testing $c(p) \leq c(q) - \delta(q, p)$ which is done in $O(1)$-time. If $p \notin S$ then we have to do the full test on emptiness of $T(q \mid S_i)$ which is done in $O(|M|)$ time.

The proof of correctness of Alg. 2 is based on the following Lemma:

Lemma 5. *Let the condition* (4) *hold. After Step 6 of Alg. 2 is performed it holds that $T(q_i \mid S_{i+1}) \subseteq S$, where S is $\sigma\delta$-stable matching on (P, c, Z).*

Proof. The set $T(q_i \mid S_{i+1})$ is $\sigma\delta$-stable if it is not influenced by pairs from L_{i+1}. Let us suppose the existence of $s \in L_{i+1}$ such that $t \in T(q_i \mid S_{i+1})$ and $t \in D(s \mid S_{i+1})$. Two conditions must hold:

$$c(t) > c(q_i) - \delta(q_i, t), \tag{5}$$

$$c(t) \leq c(s) + \sigma(t, s). \tag{6}$$

We introduce $\epsilon \geq 0$, so $c(s) = c(q_i) - \epsilon$ because $c(s) \leq c(q_i)$, which follows from the condition that $s \in L_{i+1}$. From this and from (6) we get

$$c(t) - c(q_i) \leq \sigma(t, s) - \epsilon. \tag{7}$$

From (4) we have $\sigma(t, s) + \delta(q_i, t) \leq 0$. From (7) it therefore follows that

$$c(t) - c(q_i) \leq \sigma(t, s) - \epsilon \leq -\delta(q_i, t) - \epsilon.$$

This is in contradiction with (5), since $\epsilon \geq 0$, so there are no pairs $t \in T(q_i \mid S_{i+1})$ and $s \in L_{i+1}$ such that $t \in D(s \mid S_{i+1})$ which proves the lemma. □

Theorem 2. *Alg. 2 solves the $\sigma\delta$-stable matching problem on any (P, c, Z) if* (4) *holds.*

Proof. The theorem is a corollary of Lemma 5. □

Time Complexity The time complexity of Alg. 2 is $O(|P| \cdot |M|)$, where $|P|$ is the number of initial pairs in the matching table and $|M|$ is the cardinality of quasi-stable matching found in the first step of the algorithm. If (P, c, Z) is a dense matching problem then the complexity is $O(n^3)$. It may be improved to $O(n^2 \log n)$ for X-zones, this possibility will be discussed in Section 5.1.

5 Confidently Stable Matching

Let $[\underline{c}(p), \overline{c}(p)]$ be the confidence interval for the correlation value $c(p)$ of a pair $p \in P$. A rigorous approach to constructing such intervals is given in [13]. We may require stability to hold *for the worst possible combination* of correlations from these intervals. Such matching will certainly be very little sensitive to (good) noise. It can be regarded as stable with certainty related to the confidence level chosen. In practice, this is of course limited by the capture probability of the confidence interval estimator.

In this section we define confidently stable matching and show that the problem can be reduced to $\sigma\delta$-stable matching. It will be shown that Alg. 2 can be used without any modification for one of the two possible reductions of confidently stable matching.

Definition 8. *Matching S is confidently stable if for each $p \in S$ and each $q \in Z(p)$ the following holds:*

$$\text{If } \overline{c}(q) \geq \underline{c}(p) \text{ then there must be } r \in S \text{ such that } \underline{c}(r) > \overline{c}(q). \tag{8}$$

Let $\Delta(\cdot)$ be confidence interval width. There are two possible ways to reducing confidently stable matching problem to $\sigma\delta$-stable matching problem:

1. When $\underline{c}(p) \overset{\text{def}}{=} \overline{c}(p) - \Delta(p)$ and $\underline{c}(r) \overset{\text{def}}{=} \overline{c}(r) - \Delta(r)$. Then the condition (8) reads as follows: If $\overline{c}(q) \geq \overline{c}(p) - \Delta(p)$ then there must be $r \in S$ such that $\overline{c}(r) - \Delta(r) > \overline{c}(q)$, cf. (1). That is, we set $\sigma(p, q) = \Delta(p)$ and $\delta(q, r) = -\Delta(r)$. It thus holds that $\sigma(p, q) + \delta(s, p) = 0$ for each $s, q \in P$ and that the σ, δ are associated with the tested pairs.

2. When $\overline{c}(q) \overset{\text{def}}{=} \underline{c}(q) + \Delta(q)$. Then the condition (8) is as follows: If $\underline{c}(q) + \Delta(q) \geq \underline{c}(p)$ then there must be $r \in S$ such that $\underline{c}(r) > \underline{c}(q) + \Delta(q)$. That is, we set $\sigma(p, q) = \Delta(q)$ and $\delta(q, r) = -\Delta(q)$. So it holds $\sigma(p, q) + \delta(q, r) = 0$ for each $p, r \in P$ and the σ, δ are associated with the query pairs.

We can see that Alg. 2 can be used in the first case only.

Comment To reduce computational requirements, $c(p)$ can be chosen as the estimator of the upper limit of the confidence interval. This is possible in some cases, especially when $c(p)$ is a bounded function (like normalized cross-correlation). In this case we gain a great algorithmic advantage of the first reduction over the second one: the lower limit of the confidence intervals is sufficient to compute *only* on the set M returned by Alg. 1.

5.1 Confidently X-Stable Matching Algorithm

The intersection of X-zones of any two concordant pairs has at most two members. The test for the presence of pair p in any inhibition zone $X(q)$ can be done in $O(1)$ time and so can be done the test for the emptiness of $T(q \mid S)$. In this case there is an algorithm which is faster than Alg. 2.

If we consider the first reduction of confidently stable matching to $\sigma\delta$-stable matching we obtain the following algorithm:

Algorithm 3 (Confidently X-Stable Matching).

Input: *Matching problem $(P, \overline{c}, X)$ and the procedure for computing $\underline{c}$.*

Output: *Confidently X-stable matching M.*

Procedure:

1. *Let $L := P$, $M := \emptyset$.*
2. *Label all $p \in P$ as not deleted.*
3. *Sort L in descending order with respect to c.*
4. *For each $q \in L$ do:*

(a) *if q is not labeled deleted then*
  i. $M := M \cup \{q\}$,
  ii. *compute* $\underline{c}(q) := \overline{c}(q) - \delta(q)$,
  iii. *label all* $p \in X(q)$ *as deleted.*
(b) *if q is labeled deleted and* $T(q \mid M) = \emptyset$
  i. $M := M \setminus X(q)$,
  ii. *label all* $p \in X(q)$ *as deleted.*

Comments

1. The T set in the test in Step 4b is $T(q \mid M) = \{t \mid t \in X(q) \cap M,\ \underline{c}(t) > c(q)\}$.

2. The simplification in Step 4(b)ii of Alg. 3 over the corresponding Step 6 of Alg. 2 is due to the following observation which holds for confidently stable matching:

   Observation 2. *Let equality hold in* (4)*, i.e. let* $\sigma(t,q')+\delta(q,t) = 0$ *for* $q,t,q' \in P$, $t \neq q$, $t \neq q'$ *when* $c(q) \geq c(q')$*. Then the following holds for any* $r \in P$

   $$T(r \mid S) \cup D(r \mid S) = Z(r) \cap S, \qquad (9)$$

   where M is the pseudo-stable and S is the $\sigma\delta$-stable matching on (P,c,Z).

   Proof. The statement immediately follows from the assumption and from (2) and (3). □

Time Complexity Both the deletion flag and the emptiness of $T(q \mid M)$ can be tested in $O(1)$ time, so the overall algorithm complexity is $O(|P| \log |P|)$ which is the same as the time complexity of the optimal implementation of quasi-stable matching algorithm (Alg. 1) for X-zones.

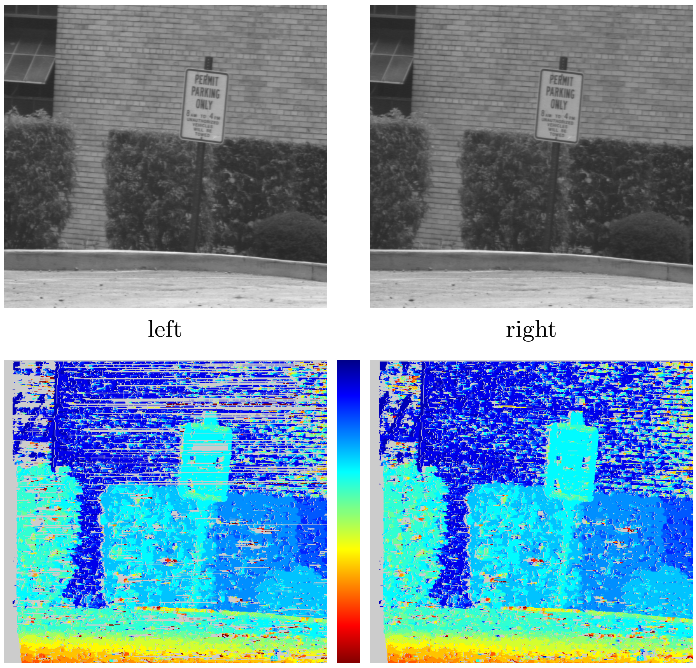
6 Experiments

6.1 Qualitative Results

In this subsection we discuss three of the most important properties of stable matchings: probability of failure, unbiasedness, and sensitivity to noise under low SNR. We will demonstrate the results on standard image sets. The next subsection then reports a ground-truth experiment.

Probability of Matching Failure Suppose that $c(p)$ is a random variable independent and identically distributed for all $p \in P$ and there is no repetition in surface texture. Under such conditions it can be shown that the probability of mismatch in FX-stable matching is a function of inhibition zone depth which increases at ever lower ratio with increasing depth if the discriminability of correlation c is greater than 0.5 [20]. Discriminability is the probability of the event that a correct match p has greater correlation $c(\cdot)$ than an incorrect match q. Discriminability of various correlation measures differs. Moreover, it typically decreases with decreasing SNR and/or decreasing matching window size. Obviously, in the presence of repetitive texture the probability of mismatch increases with inhibition zone depth. This limits the applicability of the above observation.

We will demonstrate the point on the Shrub image pair, see Fig. 3. The central object of the scene is a sign posted on a thin pole just in front of trimmed shrubbery. The leafed shrubs cannot be considered as a surface at the scale of image resolution, so no continuity prior can be used. Five individual shrubs can be distinguished in the scene, they differ by their distance from the camera. The road surface in the foreground is at high inclination angle relative to the camera pair, so in addition to the lack of texture it also exhibits large relative affine distortion between the images. The rim of the curb



left right

(a) search range $[-512, 512]$ (b) search range $[11, 30]$

Figure 3: *FX*-stable matching for the Shrub image pair. Full disparity search range (left) and search limited to the true range (right). The color bar shows how the range of disparities is color-coded to distinguish error matches more clearly (small disparities are blue, large disparities are red, unassigned disparities are gray).

is parallel to epipolar lines in the left half of the image and it is thus difficult to match. The background wall images as a very regular texture 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 reflects the sky, and thus violates the Lambertian assumption.

In Fig. 3a we can see the results of stable matching algorithm using *FX*-zone of *full depth*.[7] Modified normalized cross-correlation [15]

$$\mathrm{MNCC}(W_L, W_R) = \frac{2\,\mathrm{cov}(W_L, W_R)}{\mathrm{var}\,W_L + \mathrm{var}\,W_R} \tag{10}$$

was computed between 5×5 image windows W_L, W_R. The result in Fig. 3b was obtained with the same inhibition zone but pairs outside of the expected disparity range were removed from the matching table. Note that the matching is quite dense in Fig. 3a, even though full disparity range was searched for correspondences. This would not be possible if the probability of correct match decreased significantly with inhibition zone depth. With reduced disparity search range the density increases only slightly, mostly in areas with repetitive texture, as predicted. To conclude, *FX*-stable matching appears to have global optimality property. Note also that, unlike here, many stereo algorithms fail when disparity search range is not constrained to a small interval.

Bias and Featureless Surfaces Since stability does not explicitly assume surface continuity there is no interpolation effect over featureless areas and regions of insufficient texture remain unmatched. On the other hand, since there is no explicit continuity prior, there is no bias towards continuous surfaces. This is especially important in scenes of great depth with small (thin) objects as in Fig. 4, for instance.

We selected this image pair because of the challenges it poses to a matching algorithm. It is a scene of great depth and thin objects. The two foremost birch trees violate

[7]Full depth means the zone extends all the way to the border of matching table, which corresponds to the largest possible disparity search range.

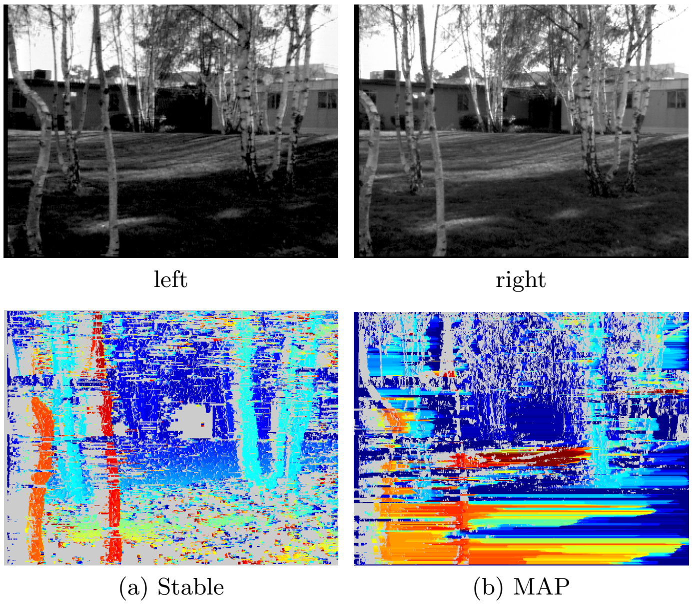


Figure 4: FX-stable matching and MAP matching via dynamic programming on the Birch Tree image pair. Disparity search range [0, 60].

ordering constraint by about their own width. The images are of rather poor quality: the shadowed floor areas are completely featureless due to under-exposure (image values are exactly zero over large areas there). So is the dark corner in the image center and the walls of the building. The sky is over-exposed and image function is again constant there. Grass and tree leafs are under-sampled. Both images differ in their mean gray-level and contrast.

In Fig. 4 we can see two disparity maps for the Birch Tree image pair: one computed as FX-stable matching and the other as Maximum Aposteriori Probability (MAP) solution using the Cox et al. algorithm based on dynamic programming [5].[8] Both algorithms use 5×5 matching window on the 484×640 images. In stable matching the thin foreground trees are mostly matched even though the (assumed) ordering does not hold because of wide stereo baseline. Because of this violation the foremost tree is broken to several sections according to which of the foreground or background texture is locally stronger. In the MAP solution the featureless ground area often receives the foreground tree disparity, which is a heavy artifact. Moreover, long sections of the foreground tree remain unmatched. We say the MAP solution is heavily biased towards continuous solutions. More on bias in Section 6.2.

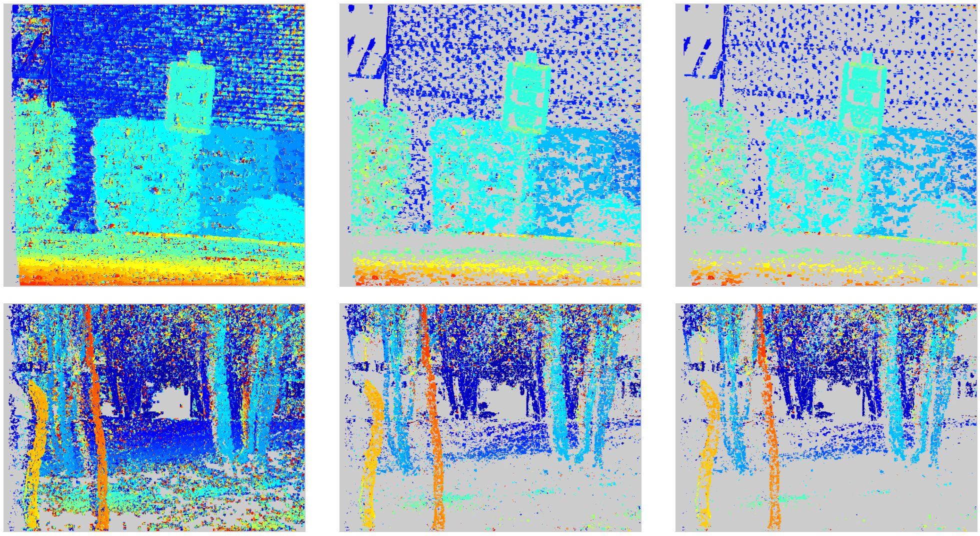
Sensitivity to noise under low SNR We can see that in featureless areas there are incorrect sparse random matches in stable matching (cf. the shadowed window area in Fig. 3 and the texture-less floor patches in Fig. 4). Confidently stable matching should clean these regions up. The results of confidently X-stable matching are shown in Fig. 5 and the results of confidently FX-stable matching are shown in Fig. 6. Alg. 3 was used for the X-stable problem and the general Alg. 2 was used for the FX-stable problem. The confidence interval was estimated as

$$\big[c(W_L, W_R) - \alpha\, \lambda(W_L, W_R),\ c(W_L, W_R)\big], \tag{11}$$

where α is a constant parameter, $c(W_L, W_R)$ is as in (10), and $\lambda(W_L, W_R)$ is derived from (10) and is as follows

$$\lambda(W_L, W_R) = \frac{4\,|c(W_L, W_R)|}{\operatorname{var} W_L + \operatorname{var} W_R}. \tag{12}$$

[8] Regularization penalty of the Cox et al. algorithm was set to 150 to achieve the visually best possible results for this particular stereo pair.



$\alpha = 0$ $\alpha = 10$ $\alpha = 20$

Figure 5: Confidently X-stable matchings for the Shrub and Birch image pairs for increasing confidence interval width, cf. (11). Results for $\alpha = 0$ are (ordinary) X-stable matchings. Disparity search range was $[10, 30]$ in Shrub pair and $[0, 60]$ in Birch pair.

The $\lambda(\cdot)$ is the sensitivity of $c(\cdot)$ to small Gaussian noise in the measurement. Note that for $\alpha = 0$ we obtain (ordinary) stable matching.

Note in Figs. 5 and 6 that widening the confidence interval results in cleaning up mismatches in low SNR areas only. This is notable in the Birch image pair, where the ground is featureless except in the illuminated areas. Mismatches are cleaned up quite well there but they mostly remain in the upper part of the image where the image texture is stronger and the confidence intervals narrow. This suggests that one cannot rely on the confidence itself, without any prior continuity model. Results with FX-stable matching are noticeably better since the X-zone of a monocularly visible pair $p \in P \setminus S$ need not contain any binocularly visible pair, whereas this is not the case in FX-stable matching.[9] Note also that for the same confidence interval the FX inhibition zone gives significantly denser disparity maps than the X zone.

6.2 Quantitative Results

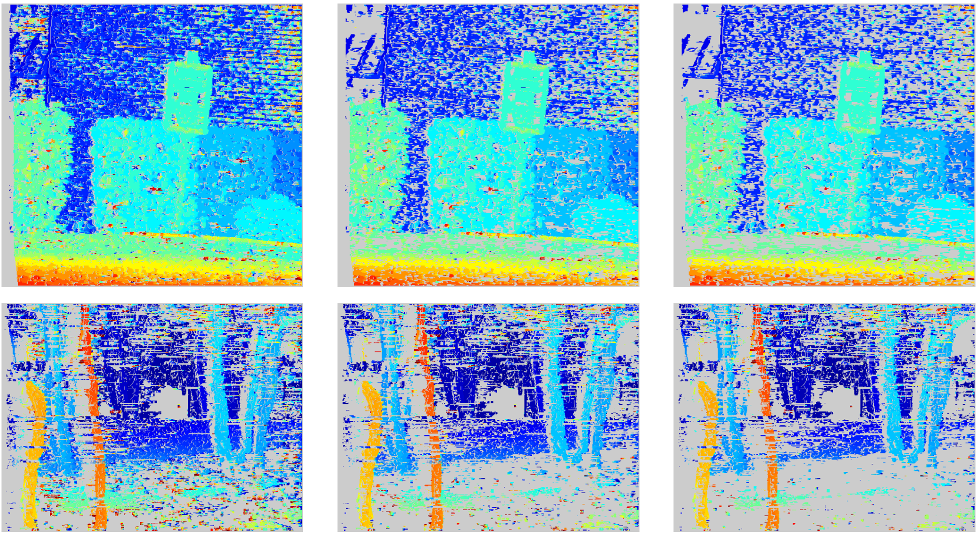
The results shown in the last subsection are not quantitative and predictive. By saying they are not predictive we mean they give us only a little hint as to the failure likelihood of any of the proposed algorithms. Here we describe experiment designed to estimate matching failure probability and bias under varying SNR.

6.2.1 Experimental Setup

The test scene we used (see Fig. 7 left) consists of three long thin textured stripes (we call it object; it is held in place by a silver U-frame visible at right) in front of a textured plane (it will be called background; it is far-right, behind the frame). Of the four cameras (mounted on a black frame at left) we used only one vertical pair in this experiment. We used digital Pulnix TM-9701 cameras, DataTranslation DT3157 frame-grabber and a custom digital multiplexer.

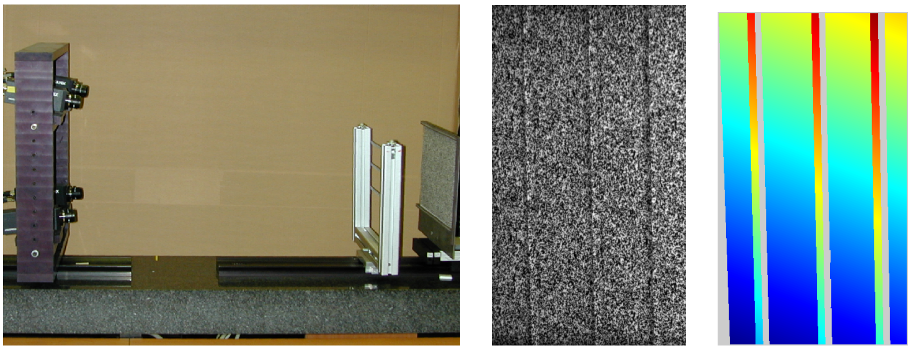
The object stripes are approximately 14 pixels wide in the images. The object-background distance was adjusted such that the width of the half-occluded region is exactly equal to the width of the stripes (widening the half-occluded region would result in violating the binocular ordering).

[9] A pair $p = (i, j)$ is monocularly visible if either $i \in I$ or $j \in J$ are half-occluded, it is binocularly visible otherwise.



$\alpha = 0$ $\alpha = 10$ $\alpha = 20$

Figure 6: Confidently *FX*-stable matchings for the Shrub and Birch image pairs for increasing confidence interval width.



experimental setup | highest-contrast left image | ground-truth disparity map

Figure 7: The test scene.

The scene was illuminated by controlled stabilized illuminant, whose adjustable intensity was used to vary texture contrast. The contrast was measured as the mean value of the left image. The smallest-contrast texture is visually hardly distinguishable from image noise (contrast value of 3.0) and the highest-contrast texture image is shown in the middle in Fig. 7 (contrast value of 74).

Ground-truth occlusion boundaries were detected manually with the help of illumination that enhanced them in the images. The ground-truth disparity within the boundaries was obtained by fitting planes to disparity map obtained from Maximum-Cost Stable Monotonic Matching algorithm [19] using the highest-contrast images. The resulting map is shown in Fig. 7 right. The cameras used in the experiment were verging towards a spatial point slightly behind the scene and their common optical axis was thus not perpendicular to the target. This results in slowly varying disparity across both the object and the background. The background disparity varies from -5 to 20 pixels and the object disparity varies from -16 to 9 pixels.

Matching All algorithms used 5×5 matching window. The MAP algorithm discussed in Section 6.1 was based on sum-of-squared-differences (SSD) correlation measure and all other algorithms used correlation (10). In all cases disparity search was done over the full range of ± 333 pixels in 587×333 images.

6.2.2 Types of Error

Five types of specific error were considered. They are all mutually related and all of them are important for assessing the quality of a matching algorithm. We assume half-occluded regions are identified.

All errors are computed from three basic matching error types:

1. *False positives*, i.e. matches found in half-occluded region,
2. *False negatives*, i.e. missing matches in binocularly viewed area (holes),
3. *Mismatches*, i.e. matches in binocularly viewed area where the difference from ground-truth was greater than 1.

In our experiment we distinguished foreground object and background object, since we are interested to see if any of the studied matchings exhibits bias towards large objects. The errors computed for the sake of the experiment were divided in three groups

1. Overall quality:

   - *Failure Rate* (FR) is the number of all mismatches and all false negatives in both the foreground object and the background normalized by the entire binocularly visible area.

     This error is related to overall matching quality but does not measure half-occluded region artifacts. It is a sum of two parts defined below:

     $$\mathrm{FR} = \frac{1}{2}(\mathrm{MR} + \mathrm{FNR}).$$

   - *False Positive Rate* (FPR) is the number of incorrectly assigned matches in monocularly visible area normalized by that area.

     This error measures the inability to correctly detect half-occluded regions.

2. Accuracy and density:

   - *Mismatch Rate* (MR) is the number of all mismatches normalized by binocularly visible area. one.

     This error measures the accuracy of matching.

   - *False Negative Rate* (FNR) is the number of missing matches (holes) in both objects normalized by the entire binocularly visible area.

     This error measures the sparsity of disparity map.

3. *Unbiasedness* (B) is the ratio of failure rates computed for the foreground and the background object independently:

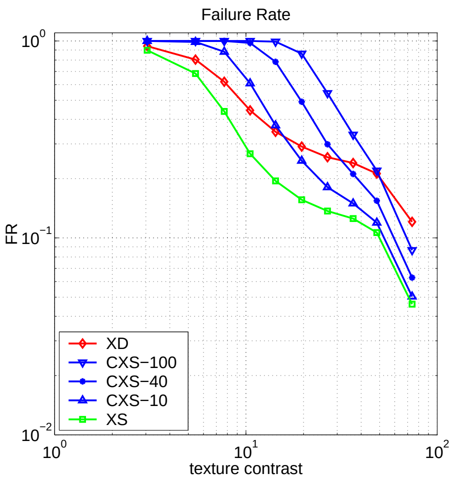
   $$B = \frac{FR_{\text{background}}}{FR_{\text{foreground}}}.$$

   Matching algorithm is biased if it assigns correct matches in large objects more often than in small objects. Small values of B imply biased matching, values $B = 1$ imply unbiased matching.

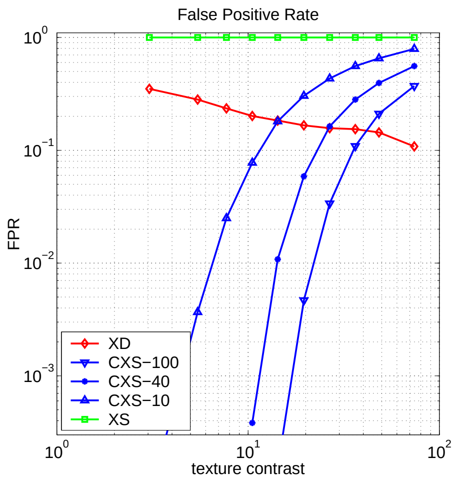
6.2.3 Discussion

The results are shown in plots in Fig. 8 for X inhibition zones and in Fig. 9 for FX inhibition zones. Texture contrast (horizontal axis) is directly related to signal-to-(quantization)-noise ratio. Vertical axes show error values. Note that both axes have logarithmic scale (except for unbiasedness). Disparity maps computed by the tested algorithms are shown in Figs. 10 and 11. The top rows give results for the maximum image contrast of 74.0 and the bottom rows give results for intermediate contrast of 10.5.

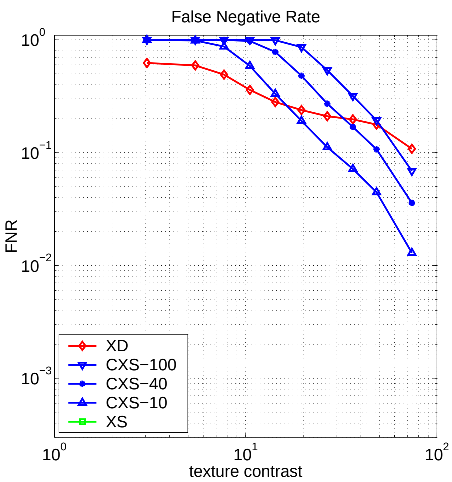
Failure Rate (see Figs. 8(a) and 9(a)) measures the overall quality of computed disparity map in binocularly visible areas. We can see the quality of FX-stable matching (FXS) is much better than in any other matching, which is confirmed visually in



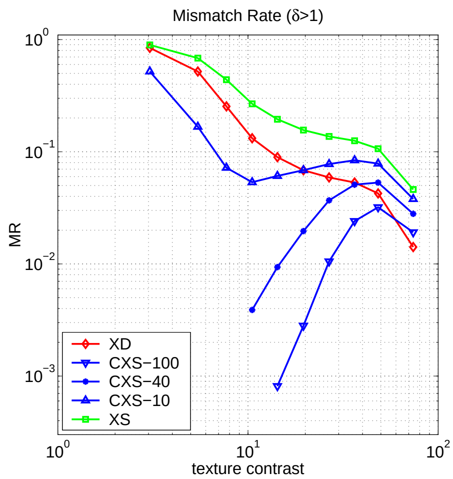
(a) binocular artifacts



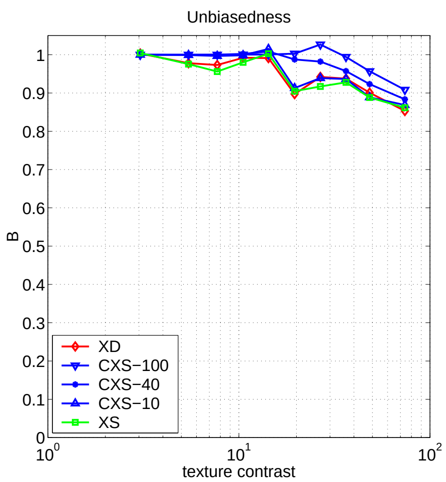
(b) monocular artifacts



(c) sparsity

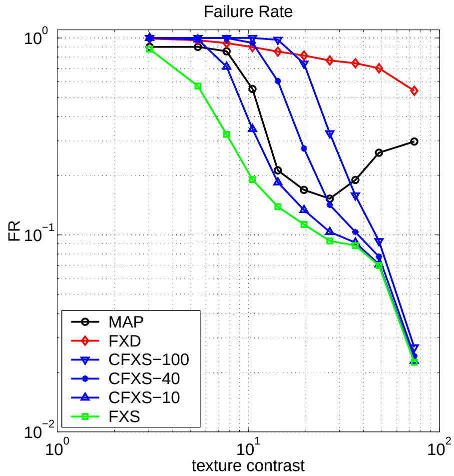


(d) inaccuracy

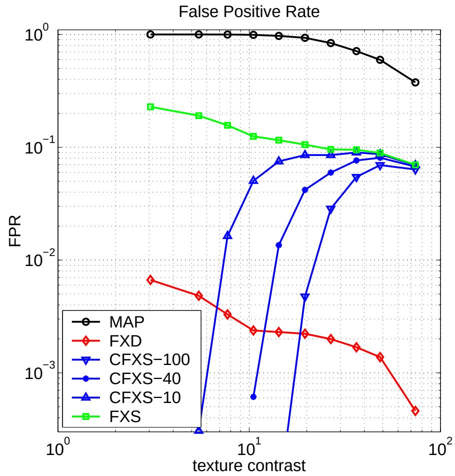


(e) unbiasedness

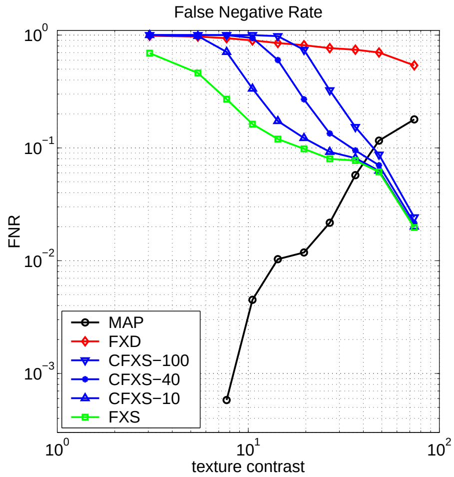
Figure 8: Five types of matching error evaluated on ground-truth test data for X-stable matchings.



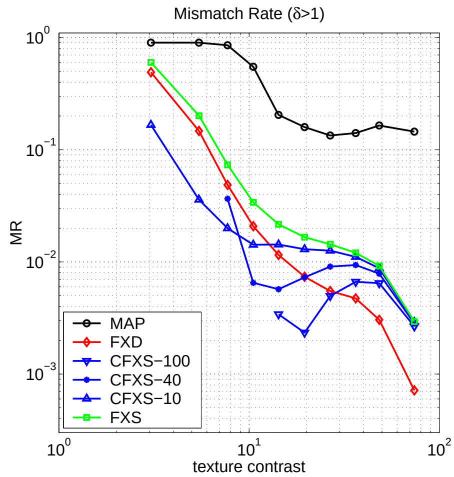
(a) binocular artifacts



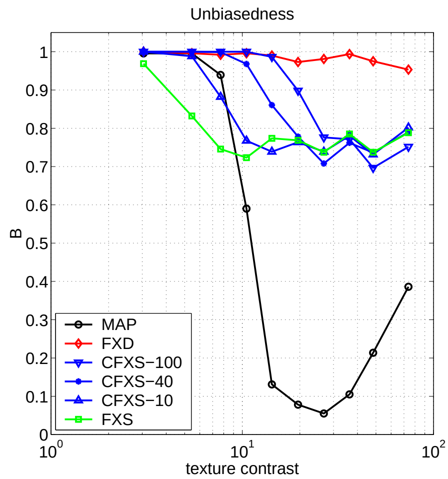
(b) monocular artifacts



(c) sparsity

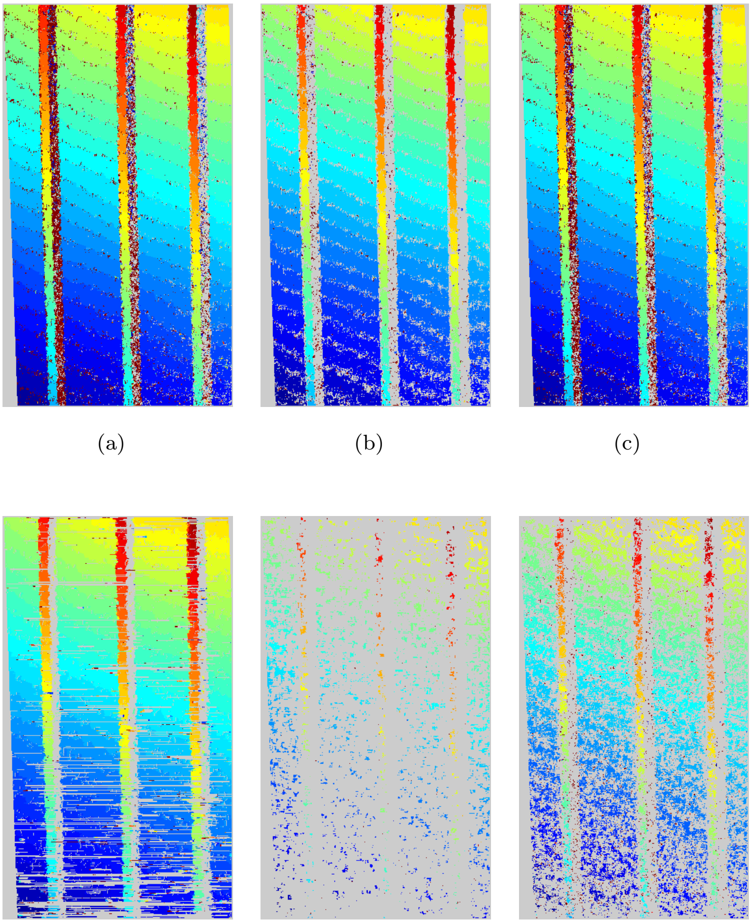


(d) inaccuracy



(e) unbiasedness

Figure 9: Five types of matching error evaluated on ground-truth test data for *FX*-stable matchings and for MAP matching.



(a) (b) (c)

(d) (e) (f)

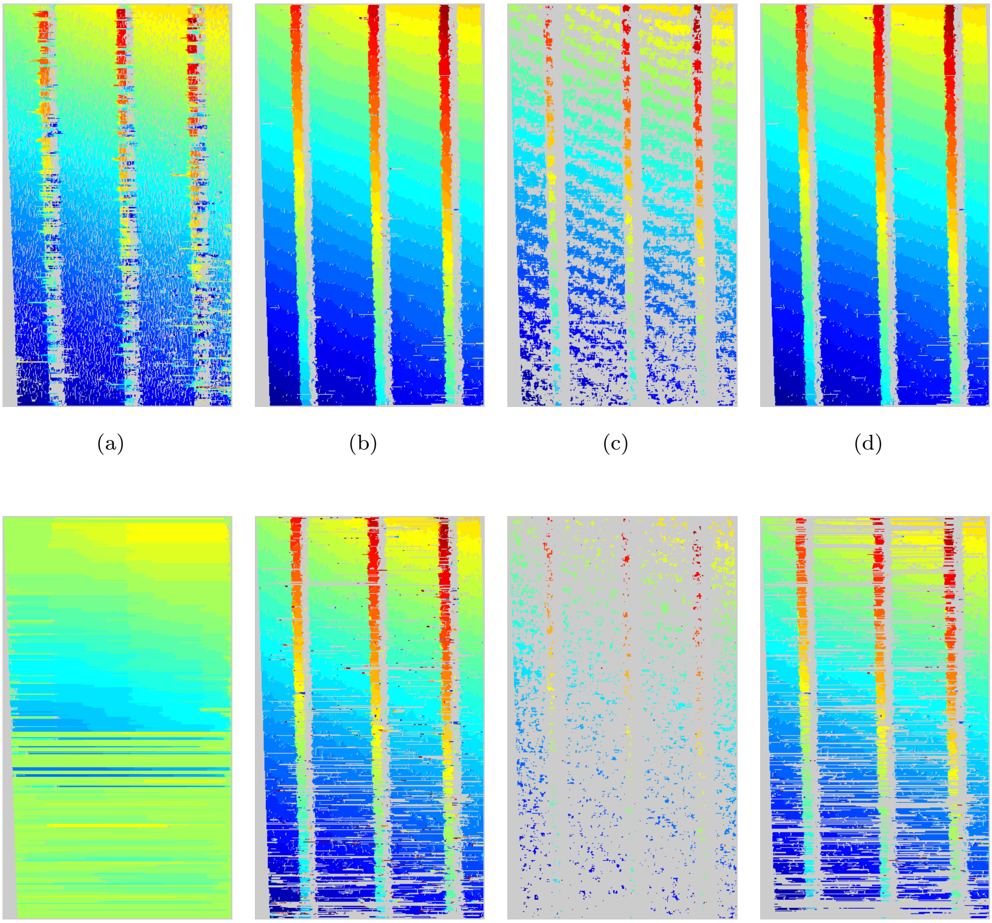
XS XD CXS-10

Figure 10: Disparity maps from algorithms based on X zone under texture contrast of 74.0 (top) and 10.5 (bottom). Unassigned matches are shown gray. The maps show the character of failure in the various tested algorithms.

Figs. 11(b) and 11(f). We can also see that there is a large difference in error rates between FX-dominant (FXD) and FX-stable matchings. Confidently stable matching (CXS, CFXS) quality depends on the parameter α; results for $\alpha = 10$ (CXS–10, CFXS–10), $\alpha = 40$ (CXS–40, CFXS–40), and $\alpha = 100$ (CXS–100, CFXS–100) are shown in Figs. 8 and 9, respectively. Increasing α decreases overall quality mostly because of increasing sparsity of the map (cf. Figs. 8(c) and 9(c)). Results of the MAP algorithm are shown for comparison. The regularization penalty of the MAP algorithm was set to the value of 500, which minimizes the false positive rate for the highest-contrast image pair. The re-ascending character of the failure rate curve is due to increasing sparsity in high-contrast images (cf. Fig. 11(a)).

False Positive Rate (see Figs. 8(b) and 9(b)) measures the inability to detect half-occluded regions. The FX-dominant matching is clearly the best for contrasts over 20. This is confirmed in Fig. 11(c). Under contrasts lower than 10 the confidently stable matchings are empty, which is why the corresponding curves start at higher values of contrast. The MAP algorithm performs the worst.

Note the increasing tendency of false positive rate in confidently stable matchings. This suggests that their ability to suppress unreliable matches weakens in high-contrast images where their performance approaches that of the ordinary stable matching. Although the ordinary X-stable matching is not able to detect half-occluded regions at all



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

(e) (f) (g) (h)

MAP FXS FXD CFXS-10

Figure 11: Disparity maps from algorithms based on *FX* zone under texture contrast of 74.0 (top row) and 10.5 (bottom row).

(it involves no model of occlusion) the results for confidently stable matching confirm it is possible to detect some half-occluded pixels based exclusively on the statistical properties of match similarity value. If geometric model of the occlusion is included (the ordering constraint), however, the additional improvement is dramatic (the FPR is by the order of magnitude better, cf. Figs. 8(b) and 9(b)). From the experiment it follows that the geometric constraint is considerably more useful for detecting half-occluded regions than the statistical properties of match correlations. For even greater accuracy (to gain more than two additional orders of magnitude) one could use confidently *FX*-dominant matching.

False Negative Rate (see Figs. 8(c) and 9(c)) measures disparity map sparsity. The sparsest maps are obtained from *FX*-dominant matching (except for small contrasts for which confident matchings are empty). The densest maps are obtained from the MAP algorithm, except for large contrasts where both *FX*-stable and confidently *FX*-stable matchings perform substantially better. The density of the MAP results is achieved at the cost of missing the foreground object completely, see Fig. 11(e).

It is interesting to note that, in false positive rate, the *FX*-dominant matching is superior to the *FX*-stable matching but in false negative rate the situation is just opposite. It seems that in stable matchings detecting objects is possible at the cost of misdetecting half-occlusions and vice versa.

Note the sparsity of *FX*-dominant matching along lines where disparity changes by a unity (Fig. 11(c)). This is probably due to sensitivity of correlation measure to image discretization. Discretization insensitive correlations [3] could help increase dominant matching density substantially.

Mismatch Rate (see Figs. 8(d) and 9(d)) measures disparity map inaccuracy. The most accurate is *FX*-dominant matching except under low contrasts where all confidently stable matchings performs better. The worst is the MAP algorithm, since it tends to either miss the foreground object or to smooth the discontinuity between the thin foreground object and the background. The re-descending character of the error in confidently stable matchings is an interesting phenomenon, probably due to rapidly increasing discriminability of correlation measure. The rapid improvement in all errors between the last two measurements is also apparent in other errors and algorithms.

Unbiasedness (see Figs. 8(e) and 9(e)) measures the ability not to miss small objects in favor of large objects (background). Methods based on strong prior continuity models will be heavily biased, methods without such models will be unbiased. We can see that the *FX*-dominant and all confidently *X*-stable matchings show no bias. The *FX*-stable matching and its confident version shows small bias towards larger objects confirming that it possesses weak global optimality property. Bias is very large in MAP algorithm except for highest-contrast images. Under the high contrast the ability of MAP to detect small objects is undermined by increased sparsity of the map, cf. Fig. 9(c) and Fig. 11(a).

Summary For high SNR the *FX*-dominant matching is the most accurate of all tested algorithms (in terms of FPR, MR). For low SNR it is the next most accurate to confidently *FX*-stable matchings. In both cases the accuracy increased at the cost of increased sparsity (FNR). Combining confidence and dominance could lead to an algorithm that finds extremely accurate (but sparse) correspondences.

FX-stable matching is the best in sparsity for high SNR (Figs. 9(c) and 11(b)) and quite good in detecting half-occlusions (the error is by about an order of magnitude smaller than in the MAP algorithm). In the overall error (FR+FPR) the *FX*-stable matching performed the best.

There is a question whether confidently X-stable matching could help in wide-baseline stereo based on interest (feature) points. Since the points are usually selected based on their distinctiveness, introducing the confidence intervals may improve the results only a little. Confidently stable matching is useful in situations when there is no prior set of interest points and we need to solve two problems at once: select distinctive points *and* solve the correspondence problem for them.

6.3 Stable Union

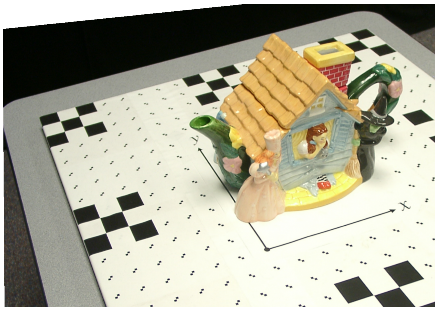
Let M_1, M_2, $\ldots, M_n$ be all subsets of P. The stable union of M_i, $i = 1, \ldots, n$ is the stable subset of their (ordinary) union. The union is computed as stable matching on the following problem

$$\left(\bigcup_{i=1}^{n} M_i,\ c^*,\ Z \right). \tag{13}$$

The correlation c^* may be constructed in several possible ways. The simplest possibility is that each pair inherits its original correlation from matching M_i whose member it is (or it retains the maximum over several correlation values if it is a member of more than just a single matching).

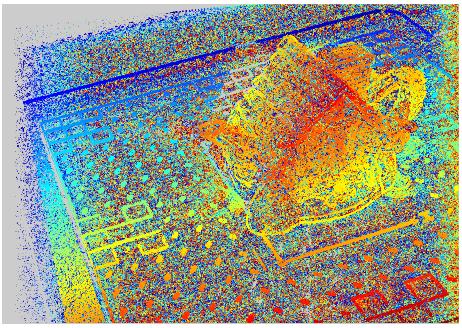
Stable unions are useful for disparity map fusion. The individual disparity maps are the matchings M_i above. The task is to produce a new matching by 'fusing' these maps. See Fig. 12 for an example of disparity map fusion. Three disparity maps from the R,G,B channels of a color image pair were obtained independently by *X*-stable matching (only two of them are shown). Note that they are quite noisy. The three maps were fused using *FX*-stable matching to obtain the final map shown in Fig. 12(f). The resulting map quality increased markedly[10] although its density is low because the glazed

[10] The goal of this example is to demonstrate the fusion principle. Normally, better results would be obtained if the correlation function c was evaluated directly on color images.

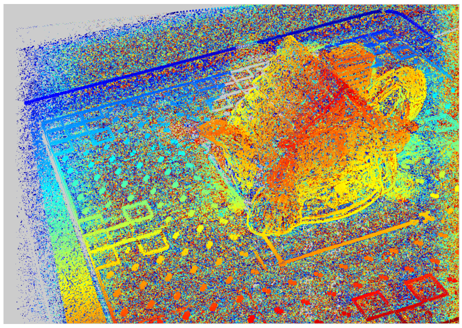


(a) left

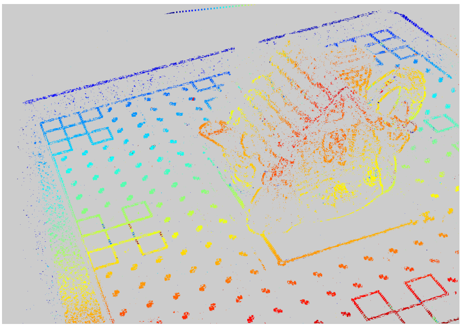
(b) right



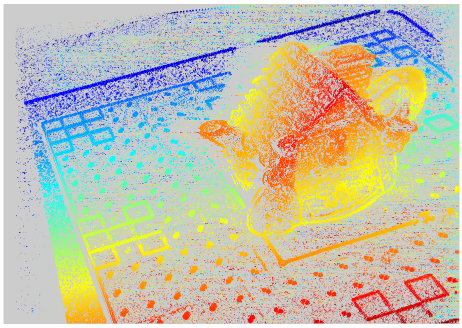
(c) R channel



(d) G channel



(e) intersection



(f) stable union

Figure 12: Fusion of disparity maps obtained independently from three channels of an RGB image. The input images (a,b), two of the three independent disparity maps (c,d), maps fused by intersection (e) and by stable union (f). Disparity map size is 1000×1500 pixels, correlation window was 5×5 pixels.

china has very little (if any) natural intrinsic texture.[11] For comparison, in Fig. 12(e) we can see the *intersection* of all the three component disparity maps. Intersection of matchings M_1, M_2, $\ldots M_n$ is the ordinary set theoretic intersection $\bigcap_{i=1}^{n} M_i$. The intersection disparity map is rather sparse. It looks quite clean except for a few gross mismatches at the borders of one of the 3×3 checkerboard patterns. Interestingly enough, the mismatches are missing in the stable union, which we attribute to weak global optimality properties of stable matchings briefly discussed in Section 6.1.

7 Discussion

It is clear that a stronger prior model results in more accurate results of any algorithm solving a vision task like stereo matching, especially under low SNR. The problem here is that a prior model which is *always* valid is too weak to be helpful in such a task (unless the application domain is strongly limited). If the prior model is strong then its violation gives rise to very strong artifacts (illusions). A good visual system must

[11] Fused matching need not be dense, or complete, because of the sparsity of population of $I \times J$ by $\bigcup M_i$.

therefore be able to reject a prior model (or to select one from a range of such models). This has been recognized as a very difficult task and the problem is still not sufficiently well understood, namely for weak prior models (like continuity). For strong models the Recover-and-Select principle has proved very successful [12]: Alternative models compete for explaining the observed data in the most compact way (a smaller number of models with fewer parameters give more compact explanation).

From the very beginning of computational stereopsis research the emphasis was put on prior models based on continuity or coherence-principle: basically a solution is better if it is continuous. This, unfortunately, results in the artifacts discussed above. The goal of this work was not to select a good prior model, our goal was to explore the case when the prior model is not available *at all.* To our knowledge, this problem has not been studied systematically in stereo vision. We believe that only after we understand this case well enough it will be possible to study the problem of model selection.

As it turns out, the solution suggested in this work opens also the way to the solution of the problem of prior model selection. This is possible since the stability principle is a fairly general concept allowing for a great versatility in defining matching problems:

A. Competition among potential matches may be defined by *other rules* than by just comparison of correlation costs and one is still able to compute stable matchings. There are two interesting cases to explore:

   1. Multi-criterial stability, which is particularly useful for multi-spectral or multichannel matching and multi-view fusion.

      It is also useful when one needs to make correspondence decisions based on similarity of neighborhood, color, local image variation, shading, photometric stereo information or photometric ratios [21], etc. In such a case one can neither construct any overall match similarity measure nor can one simply combine partial measures by majority voting or similar methods. Moreover, rules are necessary that control which information is to be considered dominant (similarity of neighborhood in highly textured areas, shading in untextured areas, etc.).

   2. The rules may be generated by a high-level vision process according to the current scene context. This opens the possibility to influence the low-level process of stereo matching by prior information.

   As long as the rules on which stable matching is based are defined on *pairs* the existence and uniqueness theorem holds. There is also a possibility to define *preferences* that rank the potential matches to J for each $i \in I$ and to I for each $j \in J$. If the preferences are not symmetric, the matching problem does not have a unique solution. This problem is called the Stable Marriage Problem [7].

B. If the *inhibition zone shape and depth are varied*, not only bipartite and monotonic matching are possible but also piecewise monotonic matching (when the F inhibition zone is of finite depth) or even non-unique matchings (when the X-zone is of finite depth) are possible. This is particularly appealing in the case when there may be multiple matches for each image pixel as in semi-transparent objects or on occluding boundaries.

In our view, non-unique matchings with finite-depth FX-zone are in fact potential match *aggregates* rather than matchings. After an aggregation step, a pair $p \in P$ may belong to one of a few (semi-)local aggregates. For each aggregate it is possible to re-compute the correlation cost much like on the disparity components in the work of Boykov et al. [4]. As opposed to Boykov et al. we are not limited to disparity components (aggregates) of equal disparity, they may even be the entire matchings. A new competition relation for pairs may then be defined on ranking provided by correlations computed on entire aggregates. The new point-wise matchings are computed from the aggregates via stable union. This scheme may then avoid the necessity of

1. using matching windows of pre-selected size, since the aggregates will play the role of the matching window, and their effective size will be data-dependent,

2. the (somewhat arbitrary) uniqueness constraint, since the uniqueness is no longer necessary to find an aggregation, and
3. the global ordering constraint, which may be quite restrictive in large-depth scenes.

These extensions are topics for our ongoing research.

8 Conclusion

Stable matchings and their confident variants show interesting properties. The dominant matchings have extremely low false positive error (in half-occluded regions). All confident matchings correctly handle the case when data do not support the correspondence hypothesis (like in featureless areas). The accuracy of confidently stable matchings does not degrade with worsening signal-to-noise ratio, only the density of matches decreases.

Reliable disparity maps can serve as initial guesses to more elaborate matching algorithms that use prior models. The task for such algorithms would be to "fill the holes." Half-occluded regions are partly identified, therefore a prior model can be employed that captures their contiguity.

Stable matchings are suitable for sparsely populated matching tables. They can be used for matching problems without additional structure (like ordering) which occur in problems like wide-baseline stereo matching. Stable matching is very efficient in fusing multiple disparity maps via stable union.

Half-occluded regions are detected based on a prior geometric model (spatial ordering) *and* unambiguous data supporting the correspondence hypothesis.

Stable matching algorithms have very few parameters: the expected disparity range (usually set to infinity), matching window size (usually 5×5 window suffices), and the confidence level in confident matchings.

Acknowledgements

This research has been supported by Grant Agency of the Czech Republic under grant GACR 102/01/1371 and by Czech Ministry of Education under grant MSM 212300013.

The standard test images were obtained courtesy Carnegie Mellon University (the VASC database) and SRI.

References

[1] S. Baker, T. Sim, and T. Kanade. A characterization of inherent stereo ambiguities. In *Proceedings 8th International Conference on Computer Vision*, volume 1, pages 428–435, Vancouver, Canada, July 2001. IEEE Computer Society Press.

[2] P. N. Belhumeur. A Bayesian approach to binocular stereopsis. *International Journal on Computer Vision*, 19(3):237–260, September 1996.

[3] S. Birchfield and C. Tomasi. A pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):401–406, April 1998.

[4] Y. Boykov, O. Veksler, and R. Zabih. Disparity component matching for visual correspondence. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 470–475, San Juan, Puerto Rico, June 1997. IEEE Computer Society Press.

[5] I. J. Cox, S. Hingorani, B. M. Maggs, and S. B. Rao. Stereo without disparity gradient smoothing: a Bayesian sensor fusion solution. In D. Hogg and R. Boyle, editors, *Proceedings of BMVC '92. British Machine Vision Conference*, pages 337–346, Leeds, UK, September 1992. Springer-Verlag.

[6] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.

[7] D. Gusfield and R. W. Irving. *The Stable Marriage Problem: Structure and Algorithms.* The MIT Press, Cambridge, London, 1989.

[8] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In H. Burhhardt and B. Neumann, editors, *Proc. European Conference on Computer Vision*, volume 1, pages 232–248, Freiburg, Germany, June 1998. Springer.

[9] B. Julesz. Towards the automation of binocular depth perception (AUTOMAP-1). In C. M. Popplewell, editor, *Proceedings of the IFIPS Congress*, Munich, 1962.

[10] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Proceedings International Conference on Computer Vision*, pages 508–515, Vancouver, Canada, July 2001.

[11] J. D. Krol and W. A. van der Grind. Rehabilitation of a classical notion of Panum's fusional area. *Perception*, 11:615–619, 1982.

[12] A. Leonardis, A. Jaklic, and F. Solina. Superquadrics for segmenting and modeling range data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1289–1295, November 1997.

[13] R. Mandelbaum, G. Kamberova, and M. Mintz. Stereo depth estimation: a confidence interval approach. In *Proceedings of IEEE 6th International Conference on Computer Vision*, pages 503–509, Bombay, January 1998. Narosa Publishing House.

[14] D. Marr. A note on the computation of binocular disparity in a symbolic, low-level visual processor. A.I. Memo 327, Artificial Intelligence Lab, Massachusetts Institute of Technology, 1974.

[15] H. P. Moravec. Towards automatic visual obstacle avoidance. In *Proc. 5th Int. Joint Conf. Artifficial Intell.*, page 584, 1977.

[16] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14:449–470, 1985.

[17] L. Robert and R. Deriche. Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In *Proc. Conf. ICIP*, pages 123–127, 1992.

[18] S. Roy and I. J. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *Proc. International Conference on Computer Vision*, pages 492–499, Bombay, India, January 1998. Narosa Publishing House.

[19] R. Šára. The class of stable matchings for computational stereo. Research Report CTU–CMP–1999–22, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, November 1999.

[20] R. Šára. Failure analysis of stable matchings. Research report, CITR, University of Auckland, In preparation.

[21] L. B. Wolff and E. Angelopoulou. 3-d stereo using photometric ratios. Technical Report 1993.10, Computer Vision Laboratory, Dept. of Computer Science, The John Hopkins University, Baltimore, MD, 1993.

[22] A. L. Yuille and T. Poggio. A generalized ordering constraint for stereo correspondence. A.I. Memo 777, Artifical Intelligence Lab, Massachusetts Institute of Technology, May 1984.

[23] C. L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):675–684, July 2000.