



A Trainable Markov Random Field for Low-Level Image Feature Matching with Spatial Relationships

TIMO DICKSCHEID & WOLFGANG FÖRSTNER, Bonn

Keywords: local features, feature matching, wide baseline stereo, image orientation

Summary: Many vision applications rely on local features for image analysis, notably in the areas of object recognition, image registration and camera calibration. One important example in photogrammetry are fully automatic algorithms for relative image orientation. Such applications rely on a matching algorithm to extract a sufficient number of correct feature correspondences at acceptable outlier rates, which is most often based on the similarity of feature descriptions. When the number of detected features is low, it is advisable to use multiple feature detectors with complementary properties. When feature similarity is not sufficient for matching, spatial feature relationships provide valuable information. In this work, a highly generic matching algorithm is proposed which is based on a trainable Markov random field (MRF). It is able to incorporate almost arbitrary combinations of features, similarity measures and pairwise spatial relationships, and has a clear statistical interpretation. A major novelty is its ability to compensate for weaknesses in one information cue by implicitly exploiting the strengths of others.

Zusammenfassung: Ein trainierbares Markoff-Zufallsfeld für die Zuordnung lokaler Bildmerkmale unter Berücksichtigung ihrer räumlichen Beziehungen. Viele Anwendungen im Bereich des maschinellen Sehens nutzen lokale Merkmale für die Bildanalyse, insbesondere in den Bereichen Objekterkennung, Bildregistrierung und Kamerakalibrierung. Ein wichtiges Beispiel in der Photogrammetrie sind vollautomatische Algorithmen für die relative Kameraorientierung. Dazu muss aus den Bildmerkmalen verschiedener Bilder anhand eines Matchingalgorithmus eine ausreichende Anzahl von Zuordnungen mit vertretbarem Ausreißeranteil gewonnen werden. Die Suche nach Zuordnungen basiert dabei meist auf der Ähnlichkeit von Merkmalsbeschreibungen. Wenn die Anzahl der extrahierten Merkmale gering ist, macht es Sinn, mehrere möglichst komplementäre Merkmalsdetektoren gleichzeitig einzusetzen. Ist die Ähnlichkeit von Bildmerkmalen kein ausreichendes Kriterium für die Zuordnung, liefern räumliche Beziehungen von Merkmalen zusätzlich wertvolle Information. In dieser Arbeit stellen wir ein allgemeines Matchingverfahren vor, das auf einem trainierbaren Markoff-Zufallsfeld basiert. Es ermöglicht die gleichzeitige Berücksichtigung nahezu beliebiger Arten von Bildmerkmalen, Ähnlichkeitsmaßen und paarweisen räumlichen Beziehungen, und lässt sich statistisch klar interpretieren. Eine Besonderheit dieses Verfahrens ist seine Eigenschaft, Schwachpunkte einer Informationsquelle durch die Stärken einer anderen implizit auszugleichen.

1 Introduction

Many vision applications use local image features as a sparse representation of image content. In photogrammetry, local image features have been used successfully to build automatic algorithms for relative image orientation (POLLEFEYS et al. 2000, MAYER 2005,

LÄBE & FÖRSTNER 2006, SNAVELY et al. 2006). In general, bundle adjustment benefits from feature matching for automatic extraction of tie points, if outliers are handled in a reliable way.

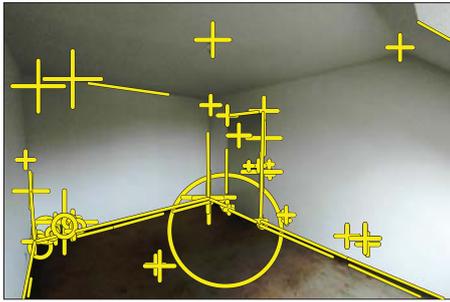


Fig. 1: Set of complementary image features covering an indoor scene. The plot includes junctions (FÖRSTNER et al. 2009) represented by crosses, blobs (LOWE 2004) represented by circles and straight line segments (FÖRSTNER 1994).

Typical image features used as tie points include corners, junctions, dark and bright blobs or line segments Fig. 1. Selecting an appropriate detector is crucial, because the importance of properties like repeatability, localization accuracy, coverage or computational complexity depends on the task at hand. If the amount of detected features is too small, it becomes necessary to combine multiple detectors. In such a case, complementarity of features is an important aspect (DICKSCHEID et al. 2010).

An image feature defines a local image patch with a particular location, orientation, shape and size. In many applications, feature correspondences can be reliably determined by analysing the similarity of those patches (*appearance-based matching*, see section 2). Feature detection and matching have to reflect the expected range of image poses. We assume the general case of arbitrary mutual rotation and possibly large scale differences, referring to all close range applications or cases where no prior knowledge on the exterior orientation is available.

Feature appearance alone is not sufficient for reliable matching if many features have similar appearance, or if the descriptors have poor distinctiveness. In such cases the use of spatial feature relationships is known to improve the matching results. For example, consider the illustration in Fig. 2. Here, the blob feature patch p_1 in the left image has very similar appearance to p'_1 in the right image if one allows for arbitrary rotations, but this is obviously not the right candidate. However, as p_1 is left of p_2 in \mathcal{I} , but

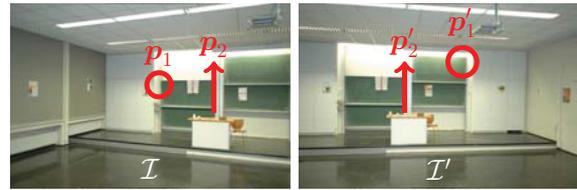


Fig. 2: Feature similarity and spatial appearance: p'_1 in image \mathcal{I}' has high similarity to p_1 in image \mathcal{I} , although it is not the right candidate. Inconsistency of the spatial relationship “is left of” with (p_2, p'_2) indicates this.

p'_1 is right of p'_2 in \mathcal{I}' , the spatial alignment gives us evidence about a possible misassignment.

A number of sophisticated methods for including spatial relationships into the matching process have been proposed, but most of them are tailored to a specific type of spatial relationship, feature type, or image data. They are therefore difficult to adapt to new matching problems.

In this work, a highly generic matching algorithm is proposed which is based on a trainable Markov random field (MRF). It is able to incorporate almost arbitrary combinations of features, feature descriptions, similarity measures and pairwise spatial relationships. The solution has a clear interpretation as the maximum a posteriori estimate of a binary classification problem, which consists in choosing a good subset from a coarse initial preselection of putative matches. A major novelty of this algorithm is its ability to compensate for weaknesses in one information cue by implicitly exploiting the strength of others.

Notation

We assume that we are given two images \mathcal{I} and \mathcal{I}' . Feature detection and description gives us two sets \mathcal{P} and \mathcal{P}' of features. Each feature $p_i \in \mathcal{P}$ is itself a set $\{(x_i, y_i), \sigma_i, \alpha_i, \mathbf{d}_i, \lambda_i\}$, where (x_i, y_i) is its location in the image given in pixels, usually referring to the centre of the local patch, σ_i is the scale given in pixels, which we assume to be proportional to the (mean) diameter of the local patch, α_i is its dominant orientation in radians, \mathbf{d}_i is the descriptor for the local patch, and λ_i denotes the type of detector used. The features also carry uncertainty information for their elements. We will discuss

this aspect in section 3.2. A feature correspondence is a pair $v_n = (\mathbf{p}_i, \mathbf{p}'_j)$ where $\mathbf{p}_i \in \mathcal{P}$ and $\mathbf{p}'_j \in \mathcal{P}'$. Dissimilarity is one property of a correspondence, expressed by a distance measure $s_n = d(\mathbf{d}_i, \mathbf{d}'_j)$ on the descriptors.

2 Appearance-Based Feature Matching

In appearance-based feature matching, one usually assumes that two features are likely to correspond if (1) the similarity of their patches is high and (2) the similarity to other patches is significantly smaller, which directly corresponds to the criteria “similarity” and “exclusion” that ULLMAN (1979) already identified for a good visual mapping. To measure similarity, robust and distinctive *descriptors* are computed from the local patches. Corresponding image features can then be found by nearest neighbour search in the space of these descriptors. A de-facto standard, often denoted as *best matching*, works as follows:

1. Determine for each descriptor in one view its two nearest neighbours belonging to the other image.
2. Select the nearest neighbour as a correspondence only if the ratio between the two distances is significantly different from 1.

A more general formulation of this algorithm leads to the BESTMATCH-K algorithm: Instead of evaluating the ratio of distances to the best and second best neighbour (BESTMATCH-2), one may consider the distances to the $(k-1)$ -th and k -th best neighbours. In the special case of $k = 1$, the nearest neighbour is always selected.

3 Spatial Relationships for Feature Matching

3.1 Related Work

In their early work, BARNARD & THOMPSON (1980) use an iterative relaxation labelling technique to select matches with locally similar image disparities. FÖRSTNER (1986) accepts only matches that are consistent under a global affine transformation, assuming that the scene can be reasonably approximated by a tilted plane. In the *relational matching* approach of (SHAPIRO & HARALICK 1987), an optimal

matching is found by minimizing the number of arbitrary spatial relationships that are not preserved by a final assignment. AGUILAR et al. (2009) proposed an iterative algorithm that constructs a consistent set of matches in terms of spatial nearest neighbourhood relationships. BAY et al. (2005) match straight line segments by first selecting the three most similar candidates per feature, and then iteratively removing matches that cause the highest number of sidedness violations (section 3.3). A remarkable feature of their algorithm is the boosting step, where previously discarded candidates are explicitly reintroduced in a post processing manner in case they become spatially consistent after the initial filtering. Most of these methods treat spatial consistency as a hard constraint, and tend to eliminate a significant amount of inliers.

To exploit appearance and spatial layout simultaneously, DELPONTE et al. (2006) exploit the properties of a *singular value decomposition (SVD)* to amplify the values of favourable matches in a *proximity matrix* $\mathbf{G} \in \mathbb{R}^{|\mathcal{P}| \times |\mathcal{P}'|}$, which captures the proximity and similarity of all pairs of features. TELL & CARLSSON (2002) proposed an interesting feature descriptor that itself captures aspects of spatial layout. Some recent methods cast spatial inconsistency and feature dissimilarity into a combined energy function to find the best matching as the one with minimum energy (SCHELLEWALD & SCHNÖRR 2005, CHOI & KWEON 2009, TORRESANI et al. 2008).

These algorithms provide no straightforward way to bring larger sets of relationships and feature dissimilarity measures with possibly significantly different strengths into such a joint formulation. Furthermore, the relative weighting of appearance and spatial consistency is not intuitive in most of these works.

3.2 Homogeneous Point and Line Representations

To derive spatial relationships for different types of features, we assume that we can always construct the normalized 2D homogeneous point $\mathbf{x}_i = [x_i, y_i, 1]^T$ with 3×3 covariance matrix Σ_{xx} representing the position of an image feature \mathbf{p}_i . For line segments, we will use the midpoint for construct-

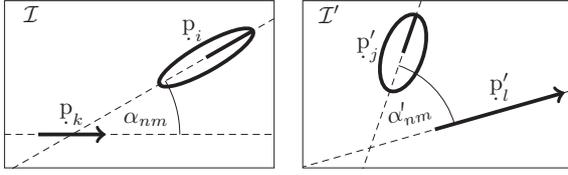


Fig. 3: Computation of the pairwise orientation difference $t_{nm}^\alpha = \min(|\alpha_{nm} - \alpha'_{nm}|, 2\pi - |\alpha_{nm} - \alpha'_{nm}|)$ for two matches $v_n = (\mathbf{p}_i, \mathbf{p}'_j)$ and $v_m = (\mathbf{p}_k, \mathbf{p}'_l)$.

ing \mathbf{x}_i , which usually has a strong localization error along the line, and a small error perpendicular to it. In a similar manner, we assume that the uncertain 2D homogeneous line $\mathbf{l}_i = \pm [\cos \alpha_i, \sin \alpha_i, -d]^\top$ with covariance matrix Σ_{ll} can be constructed from each feature \mathbf{p}_i . For point-like features, we use the centroid representation of straight line segments (MEADOW et al. 2009, 3.1.2), where the centroid is the image location of the original point feature, and the direction is identified with the dominant gradient orientation within the local patch, as stored in the SIFT descriptor. As pointed out by LOWE (2004, section 5), we must expect this direction to have a standard deviation of about three degrees. Using these conversions, we define explicit operators $\mathbf{x}(\mathbf{p}_i)$ and $\mathbf{l}(\mathbf{p}_i)$, which return the uncertain homogeneous 2D point or line representation for a feature \mathbf{p}_i .

3.3 Some Uncertain Spatial Relationships

Here we give a short description of three different kinds of relationships, leading to statistically motivated soft constraints for the matching.

Consistency of angles

Consider the angle between two oriented features Fig. 3. We assume that the difference between these angles is rather small for valid pairs of correspondences, so that large differences indicate outliers. For two putative matches $v_n = (\mathbf{p}_i, \mathbf{p}'_j)$ and $v_m = (\mathbf{p}_k, \mathbf{p}'_l)$, we compute the enclosing angles

$$\alpha_{nm} = \alpha(\mathbf{p}_k) - \alpha(\mathbf{p}_i) \pmod{2\pi} \quad (1)$$

$$\alpha'_{nm} = \alpha(\mathbf{p}'_l) - \alpha(\mathbf{p}'_j) \pmod{2\pi} \quad (2)$$

The difference $t_{nm}^\alpha \in (0, \pi)$ of the angles spanned in the two images is then given by

$$t_{nm}^\alpha = \min(|\alpha_{nm} - \alpha'_{nm}|, 2\pi - |\alpha_{nm} - \alpha'_{nm}|) \quad (3)$$

In case that the uncertainty of feature orientations varies significantly, the distances should additionally be normalized based on their standard deviations.

Consistency of distance

If two features are located close to each other in one view, we also expect their correspondences in another view to be close. This simple reasoning based on proximity was already suggested by ULLMAN (1979). We measure the distance between two feature locations, normalize it by the length of the image diagonal, and compare it to the same normalized distance of the two corresponding features in the second image, leading to the inconsistency measure

$$t_{nm}^d = t^d(v_n, v_m) = t^d(\mathbf{p}_i, \mathbf{p}'_j, \mathbf{p}_k, \mathbf{p}'_l) \quad (4)$$

$$= \frac{|\mathbf{x}(\mathbf{p}_i) - \mathbf{x}(\mathbf{p}_k)|}{\sqrt{(N_x)^2 + (N_y)^2}} - \frac{|\mathbf{x}(\mathbf{p}'_j) - \mathbf{x}(\mathbf{p}'_l)|}{\sqrt{(N'_x)^2 + (N'_y)^2}},$$

using the vertical and horizontal dimensions N_x, N_y of an image \mathcal{I} in pixels, with $t^d \in (-1, 1)$. Again, an additional benefit can be gained by replacing the Euclidean distances with the proper test statistic, i.e. by normalizing the distances using their standard deviations.

Consistency of pairwise sidedness

Consider again the example in Fig. 2. As \mathbf{p}_1 is left of \mathbf{p}_2 , while \mathbf{p}'_1 is right of \mathbf{p}'_2 , the spatial relationship “is left of” is inconsistent. This relationship is often denoted as *sidedness* or *ordering constraint*. In most existing algorithms, it is modelled as a hard constraint, based on a tolerance of a few pixels. We propose to implement a statistical test instead, which takes the orientation accuracy into account, leading to a third, binary-valued inconsistency measure

$$t_{nm}^s \in \{0, 1\} \quad (5)$$

For example, in Fig. 4, we would obtain $t_{12}^s = 0$ indicating that the sidedness relations of v_1 and v_2 are identical in both images. Accordingly,

the inconsistency of v_1 and v_3 would produce $t_{13}^s = 1$. How to determine the sidedness is explained in more detail in DICKSCHEID (2010, section 4.3.5).

4 A Trainable Markov Random Field for Feature Matching

4.1 Feature Matching as a Labelling Problem

Our goal is to find the most probable matching under a number of reasonable assumptions. Our method is based on the assumption that it is easy to obtain an initial set $\mathcal{V} = \{v_1, \dots, v_N\}$ of putative feature correspondences that contains most of the true positives. This could be the set

$$\mathcal{V}^0 = \{(\mathbf{p}_i, \mathbf{p}'_j) \mid \mathbf{p}_i \in \mathcal{P}, \mathbf{p}'_j \in \mathcal{P}', \lambda_i = \lambda'_j\} \quad (6)$$

of all correspondences between features of the same type. In practice however, a significantly smaller set $\mathcal{V} \subset \mathcal{V}^0$ can be used which still contains the majority of true correspondences. It has been shown empirically in DICKSCHEID (2010, section 2.4) that an effective way to do so is to use the BESTMATCH-K algorithm, with the matching rank k varying for each type of detector and descriptor.

Feature matching can then be interpreted as the selection of a good subset of \mathcal{V} by assigning a label l from the set $\mathcal{L} = \{0, 1\}$ to each element in \mathcal{V} . Then each element v_n of \mathcal{V} takes the role of a binary random variable defined over the set \mathcal{L} . If $l_n = 1$, we say that “match n is selected”, otherwise “match n is discarded”. For simplicity we use the notation v_n for denoting the particular event $v_n = l_n$. A labelling

$$\mathbf{l} = \mathbf{f}(\mathcal{V}) = \{v_1, \dots, v_N\} \quad (7)$$

of all variables is a *configuration*. The principle of interpreting feature matching as a labelling problem is illustrated by an artificial example in Fig. 4.

4.2 Statistical Model with Pairwise Spatial Relationships

Appearance-based matching with descriptors, as described in section 2, computes for each putative match $v_n \in \mathcal{V}$ the dissimilarity

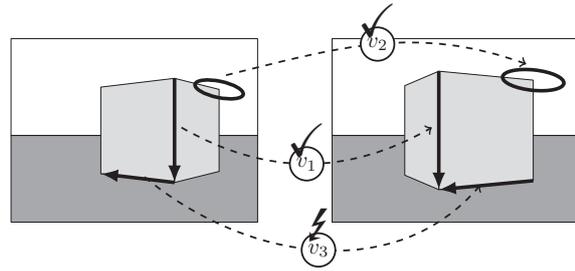


Fig. 4: Two artificial images of a scene with a cube, overlaid by three features represented by black ellipses and arrows. The features may lead to a set of three putative matches $\mathcal{V} = \{v_1, v_2, v_3\}$. The task is then to select an optimal subset of \mathcal{V} by labelling each putative match. In this example, the labelling $\mathbf{f}(\mathcal{V}) = \{v_1 = 1, v_2 = 1, v_3 = 0\}$ is the desired solution, eliminating the spatially inconsistent match v_3 .

$s_n \in \mathbb{R}$, which we collect in the vector $\mathbf{s} = [s_1, \dots, s_N]$. It then treats the decision about a match v_n without taking any spatial relations into account. In the following, we use the index set $\mathcal{N} = \{1, \dots, N\}$ over \mathcal{V} and consider pairs of correspondences defined by index pairs $\mathcal{C}_2 \subseteq \mathcal{N} \times \mathcal{N}$.

At this point, and without loss of generality, we will ignore that features and descriptors might be of different type. We will come back to this aspect again in section 5.

Considering the set $\mathcal{T}^2 = \{\mathbf{t}_{nm} \mid (n, m) \in \mathcal{C}_2\}$ of pairwise spatial inconsistencies reveals statistical dependencies between the variables. Note that $|\mathcal{T}_{nm}| = G$ is the number of different spatial relationships used. For example, the inconsistency t_{13}^s referring to “sidedness” between v_1 and v_3 is high for the two putative correspondences v_1 and v_3 in Fig. 4. We must therefore expect that one of them is an outlier, even though the descriptor dissimilarities s_1 , s_2 and s_3 might be small. In other words, after observing the spatial inconsistency, it would be naive to make independent decisions for each putative match.

In a statistical treatment, we would say that the random variable v_1 is now dependent on v_2 . It also depends on its directly related observations, s_1 and t_{12}^s . This can be expressed by an undirected graphical model as shown by the *factor graph* in Fig. 5. Each white node of this graph represents one putative feature correspondence, while shaded nodes represent ob-

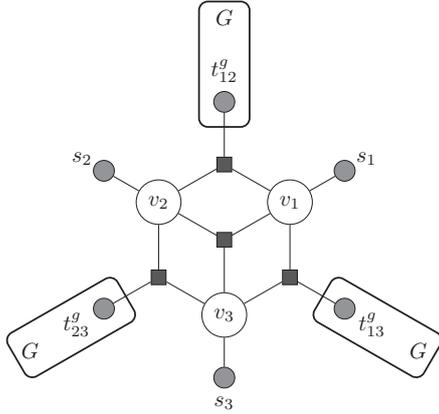


Fig. 5: Graph representing the information used for feature matching corresponding to the illustration in Fig. 4. Both descriptor dissimilarities s_n and inconsistencies t_{nm} of pairwise spatial relationships are taken into account. We follow the notation in BISHOP (2006): Observed values are represented by shaded nodes, maximum cliques (fully connected subgraphs) with more than two nodes are represented by a common black rectangle (*factor graphs*), and *plates* are used to illustrate multiple independent spatial inconsistency measures t_{nm}^g , with G being the number of different types of spatial relationships. The factor node in the centre expresses the joint prior probability for all v_n .

served dissimilarity measures. Here, values s_n refer to descriptor dissimilarities, which relate to one correspondence, while values t_{nm} refer to spatial inconsistency measures, defined over pairs of correspondences. Statistical dependencies between entities are represented by the edges of the graph. The black squares denote maximum cliques - fully connected subgraphs where all elements are conditionally dependent on each other.

In general, we make the following conditional independence assumptions:

1. All observations are mutually conditionally independent, given the correspondences.
2. The label of a putative match v_n does not depend on observations s_m or $t_{m,o}$ with $m, o \neq n$.

This model nicely supports our practical setup, as its statistical dependencies can be derived from training data.

We propose to solve the labelling problem by computing the maximum a posteriori estimate (MAP) of the variables in this model, given the observed data. Referring to the simple example

in Fig. 4, this can be done by maximizing the probability

$$\begin{aligned}
 & p(v_1, v_2, v_3, s_1, s_2, s_3, \mathbf{t}_{12}, \mathbf{t}_{13}, \mathbf{t}_{23}) \quad (8) \\
 &= P(v_1, v_2, v_3) \left[\prod_{n=1}^3 p(s_n | v_n) \right] \\
 &\quad \cdot p(\mathbf{t}_{23} | v_2, v_3) p(\mathbf{t}_{13} | v_1, v_3) p(\mathbf{t}_{12} | v_1, v_2) \\
 &= P(v_1, v_2, v_3) \left[\prod_{n=1}^3 p(s_n | v_n) \right] \\
 &\quad \cdot \prod_{g=1}^G p(t_{12}^g | v_1, v_2) p(t_{13}^g | v_1, v_3) p(t_{23}^g | v_2, v_3)
 \end{aligned}$$

The first simplification is obtained by exploiting our conditional independence assumptions, using the fact that $p(a|b, c) = p(a|b)$ in case that a is conditionally independent of c . The expansion of t_{nm} in (8) also uses the assumption that all observations are mutually independent.

For a general problem with $|\mathcal{V}| = N$ putative matches, we will obtain a graph having N binary cliques (v_n, s_n) and $G|\mathcal{C}_2|$ ternary cliques (v_n, v_m, t_{nm}^g) . The general joint probability of the variables reads

$$\begin{aligned}
 & p(\mathbf{l}, \mathbf{s}, \mathcal{T}_2) \quad (9) \\
 &= p(v_1, \dots, v_N, s_1, \dots, s_N, \mathbf{t}_{12}, \dots, \mathbf{t}_{(N-1)N}) \\
 &= P(\mathbf{l}) \left[\prod_{n \in \mathcal{N}} p(s_n | v_n) \right] \prod_{\substack{(n,m) \\ \in \mathcal{C}_2}} p(\mathbf{t}_{nm} | v_n, v_m).
 \end{aligned}$$

The factor $P(\mathbf{l})$ in (9) imposes a practical problem: It depends directly on the number of putative matches, which is unknown in advance. To make the formulation tractable for practical problems, we therefore make the following assumption for the joint probability:

$$P(\mathbf{l}) = P(v_1, \dots, v_N) \doteq \frac{1}{Z'} \prod_{\substack{(n,m) \\ \in \mathcal{C}_2}} P(v_n, v_m) \quad (10)$$

It leads to a significant simplification of the model, which now reads

$$p(\mathbf{l}, \mathbf{s}, \mathcal{T}_2) = \frac{1}{Z'} \left[\prod_{n \in \mathcal{N}} p(s_n | v_n) \right] \quad (11)$$

$$\cdot \prod_{\substack{(n,m) \\ \in \mathcal{C}_2}} p(\mathbf{t}_{nm} | v_n, v_m) P(v_n, v_m).$$

As we will see later, we do not require specific knowledge about the partition function Z' , because it does not affect the final solution.

By going from (9) to (11), we make an *explicit model assumption*. This leads to a restricted stochastic model which still corresponds to the original graphical model. The model assumption effectively drops the higher order cliques between putative matches v_n (i.e. the clique (v_1, v_2, v_3) in Fig. 5) in favour of a change of the pairwise potential functions (last factor of (11)).

4.3 Global Minimization Problem

By maximizing the density function (11) we realize a MAP estimate of the involved variables. This is equivalent to minimizing the energy function

$$E(\mathbf{l}, \mathbf{s}, \mathcal{T}_2) = - \sum_{n \in \mathcal{N}} \log p(s_n | v_n) \quad (12)$$

$$- \sum_{\substack{(n,m) \\ \in \mathcal{C}_2}} \left[\log P(v_n, v_m) + \log p(\mathbf{t}_{nm} | v_n, v_m) \right]$$

where we omit the term $1/Z'$ of the partition function, as it does not affect the solution. It is essentially a sum over functions of unary and binary cliques over \mathcal{V} , given the graphical model, and can therefore be directly interpreted as a Markov Random Field. Defining unary potentials $\theta_{n;v_n}^1$ and binary potentials $\theta_{nm;v_n v_m}^2$ as

$$\theta_{n;v_n}^1 = - \log p(s_n | v_n) \quad (13)$$

$$\theta_{nm;v_n v_m}^2 = - \log p(v_n, v_m) - \log p(\mathbf{t}_{nm} | v_n, v_m), \quad (14)$$

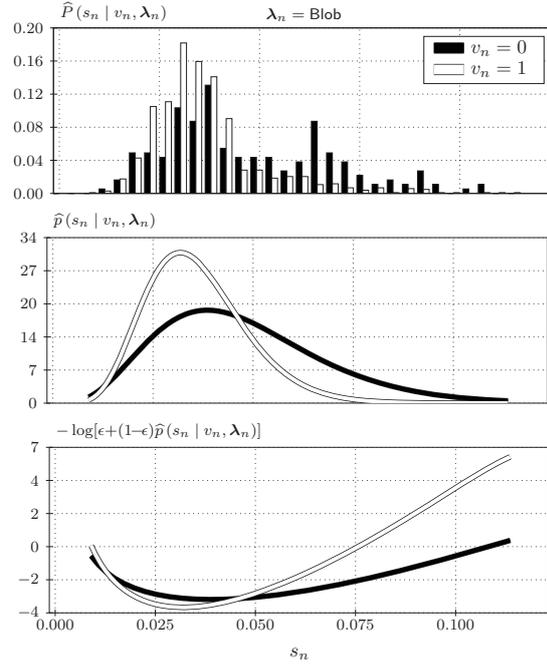


Fig. 6: *Top:* Normalized histograms of dissimilarities s_n for good ($v_n = 1$) and bad ($v_n = 0$) blob feature correspondences, computed by normalized Euclidean distances of SIFT descriptors. *Middle:* Approximation of the histograms using a Beta distribution, which is used as a parametric approximation $\hat{p}(s_n | v_n, \lambda_n)$ of the likelihood function. *Bottom:* Bounded negative log likelihood derived from $\hat{p}(s_n | v_n, \lambda_n)$, which we use for the energy potentials. The observations refer to the training dataset (section 6.1). Note that the theoretical range of the observations is $(0, 1)$, and that the Beta distribution is defined over the range $[0, 1]$. Here we only plot the relevant range; the densities are practically zero above $s_n \simeq 0.13$. The bound ϵ within the logarithm theoretically prevents the log likelihoods from reaching a limiting value for very rare values $s_n > 0.13$.

we can use the notation

$$E(f(\mathcal{V}), \mathbf{s}, \mathcal{T}_2; \boldsymbol{\theta}) \quad (15)$$

$$= \sum_{n \in \mathcal{N}} \theta_{n;v_n}^1 + \sum_{\substack{(n,m) \\ \in \mathcal{C}_2}} \theta_{nm;v_n v_m}^2,$$

which is very common in MRF theory. Note that we used $\mathbf{l} = f(\mathcal{V})$ according to (7). To find a good solution for the matching problem, given an initial set \mathcal{V} of putative matches and observations $\mathcal{D} = \{\mathbf{s}, \mathcal{T}_2\}$, we finally search for a configuration with minimum energy (12),

so we look for an optimal solution

$$f^*(\mathcal{V}) = \underset{f(\mathcal{V})}{\operatorname{argmin}} E(f(\mathcal{V}), \mathbf{s}, \mathcal{T}_2; \boldsymbol{\theta}) . \quad (16)$$

We actually apply the LP-S linear programming relaxation going back to SCHLESINGER (1976) to solve (16), which gives a very good approximation of the global optimum (KUMAR et al. 2009). We use the commercial Mosek package (<http://www.mosek.com>) for solving the relaxed minimization problem, which provides an efficient implementation of the interior point algorithm for linear programs with up to thousands of variables. The practical complexity is polynomial with very good convergence properties. For sets \mathcal{V} of putative correspondences with $N = |\mathcal{V}| < 500$, we usually solve the matching problem in a few seconds on a 2.4 GHz CPU. For sparsely textured scenes, N is typically smaller than 200, leading to negligible computation times for obtaining the optimal solution.

5 Learning the Potential Functions

Ground truth labellings for the data are obtained based on homographies (MIKOLAJCZYK et al. 2005), manually, or using the surface-based automatic annotation setup described in DICKSCHEID (2010, section 5). The latter one uses 3D point clouds from Laser scans as ground truth for the surfaces, which requires registration of the point clouds to the camera coordinate systems. The methods are indicated in the first row of Tab. 1.

For minimizing (15) one basically has to collect the potentials (13) and (14) for each node of the corresponding graph and feed them into the software. They are composed of the likelihoods of observed values, given the unknown labels and the prior probability $P(v_n, v_m)$. We will now derive some trainable parametric models for these potentials using the setup of detectors and descriptors described in section 6.1 and the spatial relationships introduced in section 3.3. We used 24 image pairs from indoor and outdoor architectural scenes for the training (section 6.2), where ground truth labellings of feature correspondences are computed using the setup described in DICKSCHEID (2010, section 5). For model fitting from ground truth data

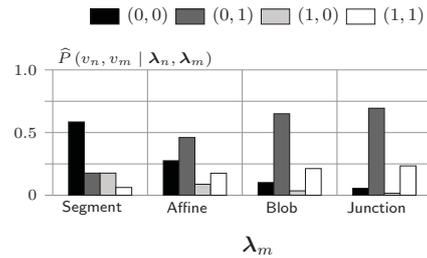


Fig. 7: Empirical fraction of pairs of putative matches, where the first match refers to straight line segments ($\lambda_n = \text{Segment}$) for different labellings l_n, l_m as observed on the training data. We obtain four groups, referring to the events $(v_n = 0, v_m = 0)$, $(v_n = 1, v_m = 0)$, $(v_n = 0, v_m = 1)$ and $(v_n = 1, v_m = 1)$. For example, if we select two matched line segments and two matched junction features from the set of putative matches, chances are around 70 percent that the line segment match is an outlier (label 0) according to the group of bars shown on the right.

we used the Statistics module of the Boost C++ Library's Math Toolkit (<http://www.boost.org>). We also tried independent datasets of the same image category and arrived at the same models with slightly different parameters. We did not investigate the dependence of the models on different image categories, so the potential functions that we present here should be considered as one particular realization of the algorithm with a focus on man-made scenes. We will evaluate this realization in section 6.

Dependency on the feature type

Our main motivation is to combine complementary feature types, descriptor dissimilarity measures with significantly different properties, and a whole range of spatial relationships simultaneously. Then the likelihoods depend formally on the type of feature λ^F , the type of descriptor λ^D , and the type of dissimilarity measure λ^M . We collect these for each putative match v_n in a tuple $\lambda_n = (\lambda_n^F, \lambda_n^D, \lambda_n^M)$. The likelihoods defining the unary and binary potentials therefore actually read $p_n(s_n|v_n, \lambda_n)$ and $p(t_{nm}|v_n, v_m, \lambda_n, \lambda_m)$. The prior becomes $P(v_n, v_m, \lambda_n, \lambda_m)$ accordingly.

5.1 Unary Potentials

Take a look at the normalized histogram on top of Fig. 6. It shows the dissimilarities of good

Tab. 1: Properties of the datasets used for our experiments.

Dataset	CLASS	BOAT	GRAFFITI	BLANK-12	BLANK-22	DRAGON
Ground Truth	manual	homography	homography	manual	manual	surface-based
Texture	sparse	strong	strong	very sparse	very sparse	sparse
3D structure	multiplanar	quasiplanar	planar	multiplanar	multiplanar	complex
Distortion	affine	rotation+scale	strong affine	affine	affine	affine
Overlap	~ 60%	~ 100%	~ 100%	~ 90%	~ 90%	~ 100%
# Images	8	6	6	12	22	6
Resolution	752 × 500	213 × 170	213 × 170	1203 × 800	752 × 500	800 × 600
Example						

($v_n = 1$) and bad ($v_n = 0$) blob feature correspondences, referring to normalized Euclidean distances of SIFT descriptors.

Due to the normalization, the histogram shapes can be reasonably approximated by a Beta distribution $\text{Beta}(s_n|a, b)$. We estimate its two parameters from training data (section 6.2) separately for the inlier and outlier distributions to obtain estimates for the class conditional likelihood functions $\hat{p}(s_n|v_n = 0, \lambda_n)$ and $\hat{p}(s_n|v_n = 1, \lambda_n)$, as shown in the middle of Fig. 6 for blob features.

The negative log likelihood $-\log \hat{p}(s_n|v_n, \lambda_n)$ that we actually use in the energy function (12) is shown in the bottom plot of Fig. 6. Note that we introduce a bound on the log likelihood by using $-\log[\epsilon + (1 - \epsilon)\hat{p}(s_n|v_n, \lambda_n)]$ with a small threshold $\epsilon = 0.001$. In practice, the bound only affects values s_n very close to the limits of the domain $[0, 1]$, which occur very rarely in practice.

We also model the dissimilarity likelihoods for other features by Beta distributions, as described in (DICKSCHEID 2010, section 4.3.3).

5.2 Binary Potentials

Priors

For each feature in one image, we preselect the k most similar features in the other image as its putative matches, where the parameter k differs between feature types λ . For example, we select more putative matches per feature for straight line segments than for blob features, following the empirical investigations in

DICKSCHEID (2010, section 2.4). We must therefore expect different prior probabilities $P(v_n, v_m, \lambda_n, \lambda_m)$. The relative frequencies within the training data for pairs of matches, where the first match is a line segment, are shown in Fig. 7. Indeed we see the strong influence of different preselection criteria per feature type on the prior: For pairs containing one line segment match and one match of another type, it is most likely that the line segment match is an outlier. This reflects the fact that k is largest for the line segments. As $P(v_n, v_m, \lambda_n, \lambda_m)$ is a discrete probability, we can model it as a binomial distribution and use those relative frequencies within the binary potentials $\theta_{nm;v_nv_m}^2$.

Likelihoods for pairs of correspondences

As an example for the likelihood $p(t_{nm}|v_n, v_m, \lambda_n, \lambda_m)$, we discuss the inconsistency t^α of angles spanned by two features (section 3.3). It is obvious that we can neither expect angles between pairs of correct matches to be always equal, nor angles between outliers to be always largely different. When investigating the empirical distribution of the consistency measures t^α on our training dataset, we see that they carry valuable information for our problem, though. The distribution for pairs of blob and junction feature matches is shown in the top row of Fig. 8. As in case of the unary potentials, we modelled the likelihoods using a Beta distribution, which corresponds strongly to the empirical distributions.

The distribution indicates that for small inconsistencies t^α between feature correspon-

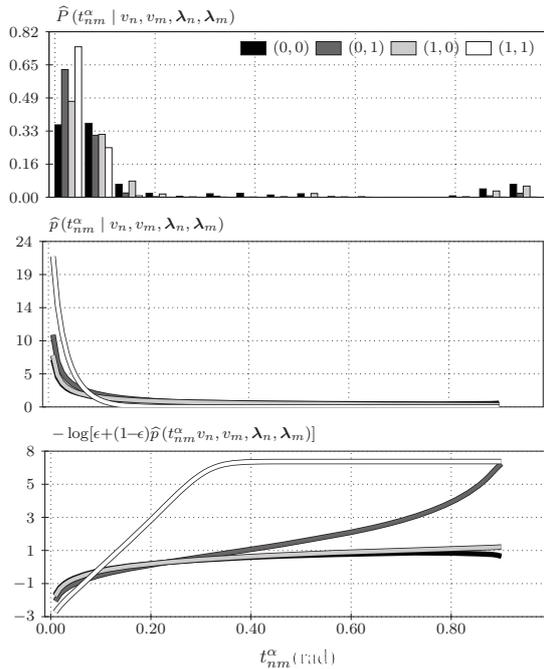


Fig. 8: *Top:* Normalized histograms of observations t_{nm}^α between blob and junction feature correspondences, denoting inconsistency of angles between pairs of oriented features. We obtain four distributions, referring to the events $(v_n = 0, v_m = 0)$, $(v_n = 1, v_m = 0)$, $(v_n = 0, v_m = 1)$ and $(v_n = 1, v_m = 1)$. *Middle:* Approximation of the histograms by Beta distributions, used as an estimate for the likelihood $p(t_{nm}^\alpha | v_n, v_m, \lambda_n, \lambda_m)$. *Bottom:* Bounded negative log likelihood derived from $\hat{p}(t_{nm}^\alpha | v_n, v_m, \lambda_n, \lambda_m)$, which we use for the energy potentials. The observations refer to the training dataset. Note that the theoretical range of the observations is $(0, 1)$, and that the Beta distribution is defined over the range $[0, 1]$. Here we only plot the range of values that we observed on the training dataset.

dences of this type, it is most likely that both matches are inliers, referring to this observation only. Hence, stronger feature types will implicitly motivate the selection of weaker ones when the angular consistency is high. With increasing inconsistency, it becomes more probable that the blob correspondence is an outlier. This corresponds strongly to our initial assumptions. For very high inconsistencies, the labelling $(0, 0)$ is motivated, which means that both correspondences are likely to be outliers.

Very similar observations can be made for other combinations of feature types. The Beta

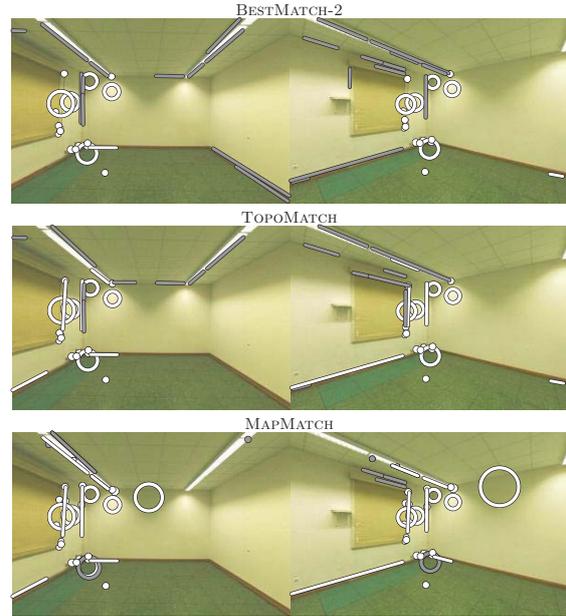


Fig. 9: Visual matching results for an image pair of the CLASS dataset for the three methods described in section 6.1. Features depicted in white are correctly matched, features in grey are outliers. We see that the simple BESTMATCH-2 approach gives quite many outliers, especially among the line segments which have the weakest descriptors. Using a topological filter and boost stage (TOPOMATCH) removes a significant number of those. The results for our approach (MAPMATCH) contain more inliers, and at the same time the lowest outlier rate. Detailed results from more image pairs of the dataset are listed in Fig. 10.

distribution is a good continuous model for all real-valued dissimilarity and inconsistency measures that we investigated, however, other measures might require a different model.

6 Experiments

We will show that the feature matching algorithm with the parametric models derived in section 5 allows for significantly better matching results on sparsely textured scenes than the standard best-matching approach (BESTMATCH-2), which only takes descriptor dissimilarities into account. We also want to make sure that our results are comparable to the results obtained with the method of (BAY et al. 2005), which is specifically designed for sparsely textured scenes. We search for a matching algorithm that maximizes the number

of correct correspondences while not exceeding a critical outlier rate. Focussing on typical image orientation algorithms with a RANSAC scheme, the critical rate is at about 50%. So, if one matching algorithm returns 20 correct correspondences with no outliers, and another one returns 50 correct correspondences with 15 outliers, we consider the latter one to be better.

6.1 Experimental Setup

Intentionally we select a set of popular feature detection algorithms with high complementarity, and use descriptors with different distinctiveness and invariance properties in order to demonstrate the potential of our method. By keeping the set of detectors and descriptors together with their parameter settings fixed, all methods shown here have to cope with the same strengths and shortcomings of the features. Nevertheless, our experiments must not be understood as a comparison of detectors, but as a comparison of wide baseline matching methods.

Detectors

The classical LOWE blob detector (LOWE 2004) is based on the Laplacian and known to have very good scale and rotation invariance. We use the original implementation kindly provided by the author, starting with the original instead of the double image resolution for building the scale space pyramid. The FOP0 detector extracts junction points using the framework of FÖRSTNER (1994). These features are not scale invariant, and also sensitive to affine distortions. We use the original implementation of the author, with a manually determined but fixed estimate of 0.015% for the standard deviation of the image noise, related to an intensity range of 1. The MSER detector of MATAS et al. (2004) detects segmentation regions with complex shape. We use the widely used implementation provided by MIKOLAJCZYK et al. (2005), however approximate the local patches by a circular shape covering an equally sized area around the same centroid. As the affine invariance that MSER itself is able to produce gets lost hereby, we denote the features as MSER^o instead. The EDGE detector from the framework of FÖRSTNER (1994) is included as a typical straight line segment detector.

Descriptors

For all but the EDGE features, we use SIFT descriptors computed using the original software provided by LOWE (2004). The feature orientations are taken from the dominant gradient orientation that is assigned to the descriptor. Descriptors for the FOP0 points are computed with a fixed window size of $3s = 12(\text{pixel})$. The straight line segments are coupled with our own implementation of the colour-histogram based descriptors of BAY et al. (2005). *These descriptors are significantly less distinctive than the SIFT descriptors.* Our implementation has been carefully compared to the implementation of the authors and leads to very similar results. Following BAY et al. (2005), the orientation of the line segments is defined by choosing the side with brighter image intensities to be left of the segment.

Matching Strategies

We show results for three different wide baseline stereo matching algorithms. The simplest and most common one is a classical descriptor-based best matching approach (BESTMATCH-2) with a 70% threshold, as described in section 2. Furthermore, we use a reimplement of the method proposed by BAY et al. (2005), which will be denoted as TOPOMATCH in the following. It includes both the three-point- and the point-line topological filtering stages described in the paper, as well as the boosting step. *Although we reimplemented the method carefully, we cannot claim that the results apply directly to the original implementation of the authors.* Our own method is denoted as MAPMATCH in the following.

6.2 Image Data

For training the likelihood functions and priors, we use observations measured from 24 pairs of images showing indoor and outdoor architectural scenes, which remain constant across all experiments. The training images are not part of the evaluation, i.e. we do not perform image specific training.

We show results based on five different datasets. The properties of the datasets are summarized in Tab. 1. The CLASS, BLANK-12

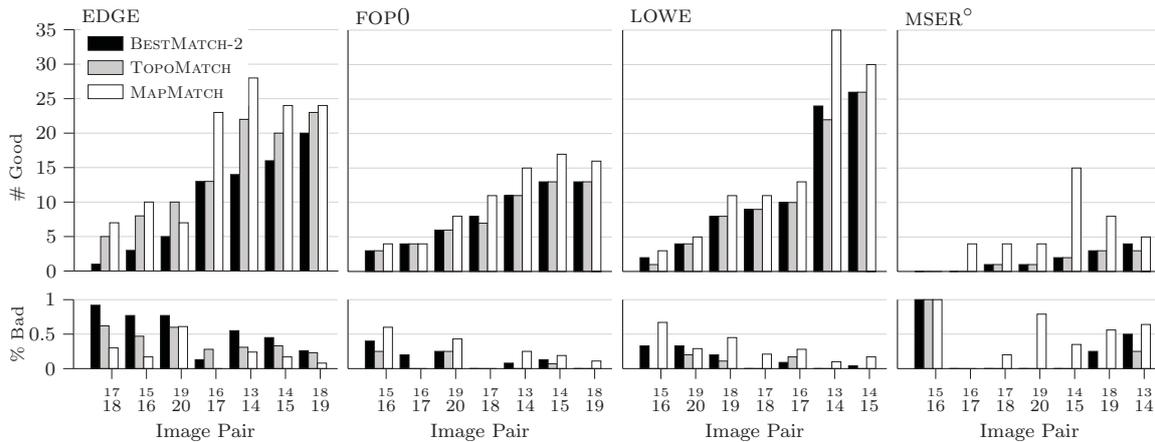


Fig. 10: Matching results for all neighbouring image pairs of the CLASS dataset, computed with the three wide baseline stereo matching algorithms described in section 6.1. Shown are the number of correct correspondences and the percentage of outliers for each feature type. The annotation has been done manually. We see that our approach (MAPMATCH) most often yields higher numbers of inliers than the others at slightly higher but acceptable outlier rates.

and BLANK-22 datasets used a fisheye lens, and have been corrected for radial distortion. As the complexity of the TOPOMATCH and MAPMATCH methods is too high for processing high resolution images with strong texture, we have downsampled the BOAT and GRAFFITI datasets (MIKOLAJCZYK et al. 2005) to reduce the amount of features in a mostly natural way.

6.3 Results

For investigating the success of a method referring directly to the extracted feature correspondences, we report the number of good correspondences (inliers) and the percentage of outliers for each matched image pair. Although we report the statistics separately for each feature type, the matching has been performed on all feature types simultaneously.

Sparsely textured datasets

Referring to the datasets with sparse texture, our approach MAPMATCH shows mostly superior matching results. The image pair of the CLASS dataset depicted in Fig. 9 provides a visual impression of the matching results for the different methods. The BESTMATCH-2 approach, relying only on descriptors, cannot compensate the weakness of the line segment descriptors, which results in many outliers among the line segment correspondences. Using the topological filter in the TOPOMATCH

method filters many of those outliers, but does not lead to a higher number of point feature correspondences. The MAPMATCH approach (bottom) achieves both effects quite well.

Fig. 10 shows detailed results for more image pairs of the CLASS dataset. Our approach yields a consistently higher number of inliers. The results for straight line segments are especially notable, as our algorithm also produces the overall smallest outlier rates. For other feature types however, it tends to have higher outlier rates than other methods.

For the BLANK-12 dataset Fig. 11, one obtains similar observations. The number of inliers is significantly higher for MAPMATCH over all considered image pairs and feature types, sometimes even better than for the other two methods. In particular, MAPMATCH would allow to compute the epipolar geometry of the third pair 6/9 quite robustly, with a total of 36 correct point matches (ignoring the line segments), while TOPOMATCH with 6 point matches is clearly at the borderline, and BESTMATCH-2 with 21 point matches significantly weaker. The TOPOMATCH implementation does not yield significantly more inliers than BESTMATCH-2, but has lower outlier rates. This is intuitive, considering that it removes matches with inconsistent spatial relationships.

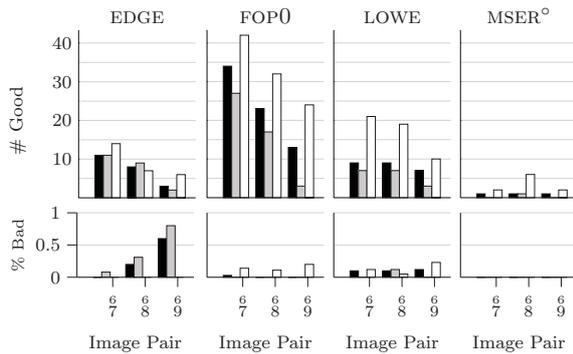


Fig. 11: Results for three image pairs with increasing baseline taken from the BLANK-12 dataset. The number of inliers is significantly higher for MAPMATCH, while the outlier rates are still good, sometimes also better than for the other two methods. In particular, MAPMATCH would allow to compute the epipolar geometry of the third pair 6/9 quite robustly, with a total of 36 correct point matches (ignoring the line segments), while TOPOMATCH with 6 point matches is clearly at the borderline, and BESTMATCH-2 with 21 point matches significantly weaker.

Standard datasets

The results for the BOAT dataset Fig. 12 show that our approach yields comparable results to the classical BESTMATCH-2. Note that here the image pairs are sorted by increasing scale and rotation difference between the images. For strong distortions, MAPMATCH yields more inliers than the BESTMATCH-2 approach, at the cost of a slightly higher outlier rate. Nevertheless it has a tendency to extract too many outliers at times, as can be seen in case of the affine region features for image pairs 1/4 in Fig. 12, and in case of the blobs for pair 1/6. The TOPOMATCH approach yields very similar results to BESTMATCH-2, with a tendency to extract even less matches. Note that although the line segments were used for matching in all of our experiments, they are not listed for the GRAFFITI and BOAT dataset, as the homography-based annotation cannot evaluate them automatically.

Results for straight line segments

The straight line features play a special role, as the matching of lines is in general more difficult due to the uncertainty of the location of the start-/endpoints, and in particular more difficult due to the weak descriptors used here.

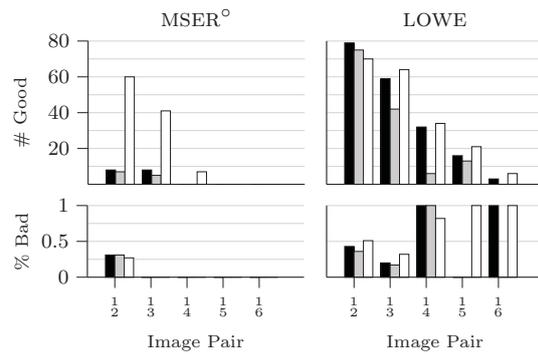


Fig. 12: Matching results for all image pairs containing the first image of the BOAT dataset, computed with the three wide baseline stereo matching algorithms described in section 6.1. The annotation has been done based on plane homographies, which works only for point features. The image scale and rotation difference per image pairs increases significantly from left to right. Due to the strong scale differences, the number of FOP0 correspondences is so small for all methods that we don't show them here.

On the investigated datasets, the MAPMATCH approach shows better results than both other methods referring to the line segments. At the same time, the TOPOMATCH method often shows better results for matching lines than BESTMATCH-2. We can therefore conclude that the spatial relationships seem to play indeed an important role for matching features with weak descriptors.

7 Conclusions and Future Work

It can be reasonable to use multiple complementary detectors in a vision system for increasing the amount of detected features. If the distinctiveness of feature descriptors is rather weak, as in case of sparsely textured scenes, or simple descriptors, spatial relationships provide important additional information for matching.

We have developed and described a generic method for modelling the matching problem with different detectors, descriptors and pairwise spatial relationships, which takes the structure of a binary classification problem and is consistent with MRF theory. Its potential functions have a clear statistical interpretation and can be trained from data using simple parametric models. The method therefore adapts

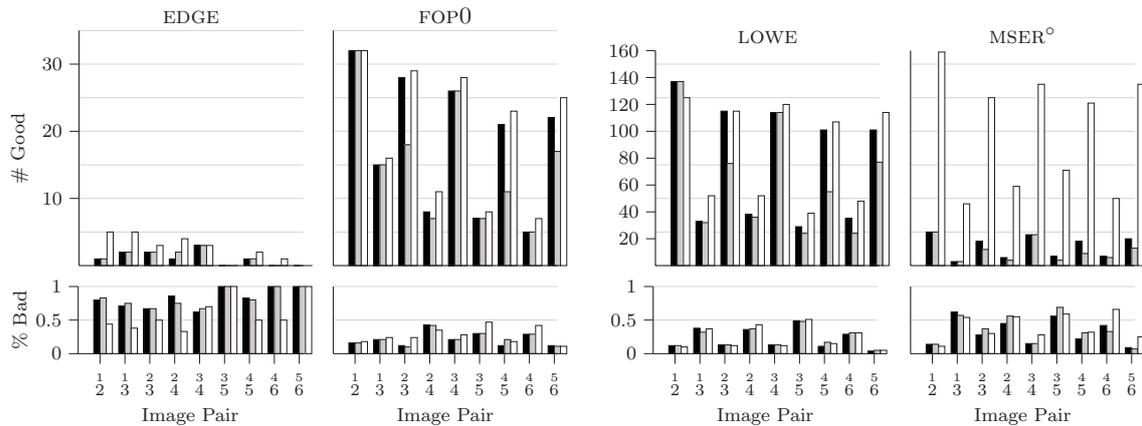


Fig. 13: Results for overlapping image pairs for the DRAGON dataset. The matching of EDGE features seems to be particularly difficult here for all three methods. The MAPMATCH approach solves it significantly better, though still not satisfyingly. For the other feature types, the MAPMATCH approach shows consistently better results in terms of higher number of inliers at comparable and satisfying outlier rates. Observe especially the affine blobs, where MAPMATCH extracts between 7 and 10 times more inliers, at a only slightly higher outlier rate.

very well to new matching problems and is straightforward to implement. We have implemented a particular instance of the algorithm which is able to produce better matching results on sparsely textured scenes compared to existing standard and specialized methods.

Simple descriptor-based matching is faster than our approach and still effective in case that many features are available. Therefore, we propose to fall back to this standard method when the amount of detected features is high.

Preselecting putative matches based on descriptor similarity is the most heuristic part of the proposed algorithm, so an investigation of more sophisticated criteria would be interesting. One may also achieve better results when choosing closer parametric approximations of the empirical likelihood distributions, perhaps by introducing mixture models. Finally, selecting more specific training images for particular matching problems can potentially give more accurate results. We have not yet investigated the effect of different training sets.

References

- AGUILAR, W., FRAUEL, Y., ESCOLANO, F., MARTINEZ-PEREZ, M., ESPINOSA-ROMERO, A. & LOZANO, M., 2009: A Robust Graph Transformation Matching for Non-Rigid Registration. – *Image and Vision Computing* **27** (7): 897–910.
- BARNARD, S.T. & THOMPSON, W.B., 1980: Disparity Analysis of Images. – *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- BAY, H., FERRARI, V. & GOOL, L.V., 2005: Wide-Baseline Stereo Matching with Line Segments. – *IEEE Conference on Computer Vision and Pattern Recognition* **1**: 329–336, Washington DC, USA.
- BISHOP, C.M., 2006: *Pattern Recognition and Machine Learning*. – Information Science and Statistics, Springer.
- CHOI, O. & KWEON, I.S., 2009: Robust feature point matching by preserving local geometric consistency. – *Computer Vision and Image Understanding* **113** (6): 726–742.
- DELPONTE, E., ISGRÒ, F., ODOONE, F. & VERRI, A., 2006: SVD-matching using SIFT features. – *Graphical models* **68** (5-6): 415–431.
- DICKSCHEID, T., 2010: *Robust Wide-Baseline Stereo Matching for Sparsely Textured Scenes*. – PhD thesis, Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn.
- DICKSCHEID, T., SCHINDLER, F. & FÖRSTNER, W., 2010: Coding Images with Local Features. – *International Journal of Computer Vision* **1–21**: 10.1007/s11263-010-0340-z.
- FÖRSTNER, W., 1986: A feature based correspondence algorithm for image matching. – *International Archives of the Photogrammetry* **26-3/3**, ISP Commission III, Rovaniemi, Finland.
- FÖRSTNER, W., 1994: A Framework for Low Level Feature Extraction. – *European Conference on Computer Vision*, Volume III: 383–394, Stockholm, Sweden.

- FÖRSTNER, W., DICKSCHEID, T. & SCHINDLER, F., 2009: Detecting Interpretable and Accurate Scale-Invariant Keypoints. – IEEE International Conference on Computer Vision, Kyoto, Japan.
- KUMAR, M.P., KOLMOGOROV, V. & TORR, P.H., 2009: An analysis of convex relaxations for map estimation of discrete mrfs. – *Journal of Machine Learning Research* **10**: 71–106.
- LÄBE, T. & FÖRSTNER, W., 2006: Automatic Relative Orientation of Images. – 5th Turkish-German Joint Geodetic Days, Berlin.
- LOWE, D.G., 2004: Distinctive Image Features from Scale-Invariant Keypoints. – *International Journal of Computer Vision* **60** (2): 91–110.
- MATAS, J., CHUM, O., URBAN, M. & PAJDLA, T., 2004: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. – *Image and Vision Computing* **22**: 761–767.
- MAYER, H., 2005: Robust Least-Squares Adjustment Based Orientation and Auto-Calibration of Wide-Baseline Image Sequences. – ISPRS Workshop BenCOS 2005: 11–17, Beijing, China.
- MEADOW, J., BEDER, C. & FÖRSTNER, W., 2009: Reasoning with uncertain points, straight lines, and straight line segments in 2d. – *ISPRS Journal of Photogrammetry and Remote Sensing* **64** (2): 125–139.
- MIKOLAJCZYK, K., TUYTELAARS, T., SCHMID, C., ZISSERMAN, A., MATAS, J., SCHAFFALITZKY, F., KADIR, T. & GOOL, L.V., 2005: A Comparison of Affine Region Detectors. – *International Journal of Computer Vision* **65** (1/2): 43–72.
- POLLEFEYS, M., KOCH, R., VERGAUWEN, M. & VAN GOOL, L., 2000: Automated Reconstruction of 3D Scenes from Sequences of Images. – *ISPRS Journal Of Photogrammetry And Remote Sensing* **55** (4): 251–267.
- SHELLEWALD, C. & SCHNÖRR, C., 2005: Probabilistic Subgraph Matching Based on Convex Relaxation. – *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCPVPR'05)* **3757**: 171–186, Springer.
- SCHLESINGER, M., 1976: Sintaksicheskiy analiz dvumernykh zritelnykh signalov v usloviyakh pomekh (Syntactic Analysis of Two-Dimensional Visual Signals in Noisy Conditions). – *Kibernetika* **4**: 113–130.
- SHAPIRO, L.G. & HARALICK, R.M., 1987: Relational matching. – *Applied Optics* **26** (10): 1845–1851.
- SNAVELY, N., SEITZ, S.M. & SZELISKI, R., 2006: Photo tourism: exploring photo collections in 3d. – *ACM SIGGRAPH 2006*: 835–846, New York, NY, USA.
- TELL, D. & CARLSSON, S., 2002: Combining Appearance and Topology for Wide Baseline Matching. – *European Conference on Computer Vision, Copenhagen*: 68–81, Springer.
- TORRESANI, L., KOLMOGOROV, V. & ROTHER, C., 2008: Feature correspondence via graph matching: Models and global optimization. – *European Conference on Computer Vision*: 596–609, Springer.
- ULLMAN, S., 1979: The interpretation of visual motion. – MIT Press, Cambridge, MA.

Address of the Authors:

Dr.-Ing. TIMO DICKSCHEID, Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, D-56068 Koblenz, Tel.: +49-261-1306-0, Fax: +49-261-1306-5302, e-mail: dickscheid@bafg.de
Prof. Dr.-Ing. Dr. h.c. mult. WOLFGANG FÖRSTNER, Rheinische Friedrich-Wilhelms-Universität zu Bonn, Institut für Geodäsie und Geoinformation, Photogrammetrie, Nußallee 15, D-53115 Bonn, Tel.: +49-228-73-2713, Fax: +49-228-73-2712, e-mail: wf@ipb.uni-bonn.de

Manuskript eingereicht: Januar 2013

Angenommen: April 2013