Virtual Line Descriptor and Semi-Local Matching Method for Reliable Feature Correspondence

Zhe Liu
zhe.liu@enpc.fr
Renaud Marlet
renaud.marlet@enpc.fr

University Paris-Est, LIGM (UMR CNRS),
Center for Visual Computing,
Ecole des Ponts ParisTech,
6-8 av. Blaise Pascal, 77455 Marne-la-Vallée, France

Abstract

Finding reliable correspondences between sets of feature points in two images remains challenging in case of ambiguities or strong transformations. In this paper, we define a photometric descriptor for virtual lines that join neighbouring feature points. We show that it can be used in the second-order term of existing graph matchers to significantly improve their accuracy. We also define a semi-local matching method based on this descriptor. We show that it is robust to strong transformations and more accurate than existing graph matchers for scenes with significant occlusions, including for very low inlier rates. Used as a preprocessor to filter outliers from match candidates, it significantly improves the robustness of RANSAC and reduces camera calibration errors.

1 Introduction

Finding reliable correspondences between sets of feature points in two images is a key step in many computer vision problems, e.g., image registration, camera calibration and object recognition. To this end, feature detectors such as SIFT [18], SURF [2], Harris-affine [20] or MSER [19] robustly identify interest points or areas in images. By design, the detected points or areas are salient enough to likely be also salient in other views of the same scene, under different imaging conditions (viewpoint, lighting, orientation, scale, etc.). Besides, these points or areas can be individually described based on their scale, if any, as well as on an abstraction of their photometric neighborhood, e.g., based on the distribution of local gradients. Such feature descriptors include SIFT [18], SURF [2] and MSER shape descriptor [12]. Like detectors, these descriptors are designed to be robust, to some extent, to variations such as noise or change of viewpoint, orientation or illumination.

Matching detected features in two images based on the similarity of their descriptor often provides good correspondences (inliers). However, it often includes false matches too (outliers). Eliminating those false matches while preserving true correspondences remains challenging for images with ambiguities or strong transformations. Ambiguity usually arises from repetitive patterns (e.g., facade windows) or lack of texture. In this case, the descriptors are not discriminative enough to safely differentiate feature points. There actually is a balance to find as repeatable descriptors tend to be less distinctive, and vice versa. As for strong transformations, they can sometimes be avoided by carefully controlling imaging conditions. Yet some sharp transformations cannot be escaped, e.g., due to strong occlusions,

© 2012. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

http://dx.doi.org/10.5244/C.26.16

when a foreground object obstructs very different background areas. To get both a high number of matching inliers and a low outlier ratio, just comparing individual feature descriptors is not enough. Global methods are required, such as RANSAC or graph matching.

Feature matching by RANSAC. For rigid transformations, RANSAC-like methods [11] can accurately separate inliers from outliers. They randomly sample subsets of correspondences to build a putative model of the transformation (fundamental/homography matrix) and count the number of matches that are compatible with the model. The largest consensus set defines what is to be considered as inliers, other matches being regarded as outliers.

This works well if the inlier rate ρ is high, not if it is low. The reason is that the number of required sampling iterations is on the order of $1/\rho^n$, where n is the number of correspondences to draw to define a model. (In general, for the fundamental matrix, $n \geq 7$ or 8 [14].) Better drawing strategies such as MLESAC [26] or PROSAC [6] can greatly reduce the number of models to sample, but they are nonetheless not well suited for inlier rates lower than 50%. Only a few methods such as ORSA [22] can treat an inlier rate of 10%. Yet in any case, all RANSAC-like methods inherently suffer from a limitation when estimating the fundamental matrix: they cannot eliminate outliers corresponding to points that have matches near their epipolar line but far from the correct location, which may degrade precision.

Graph matching methods. Graph matching is another tool to determine feature point correspondences, with a global consistency criterion. It applies not only to rigid scenes but also to deformable objects. The idea is to construct a graph where vertices are feature points and edges are pairwise relations. Higher order constraints, involving more than two vertices, can be modeled as hyperedges. Graph matching methods try to establish a vertex correspondence between two graphs, satisfying matching constraints or optimizing a global score. Some can also handle inexact matching, allowing different structures to some extent [8].

For 2nd-order graph matching, many methods use the relative distance between points as constraint [5, 16], possibly in combination with angles [3]. Feature orientation and scale are used too, e.g., to define an affine transformation predicting the projection of neighboring points [1]. Some robust pairwise descriptors combine individual feature descriptors too [13].

A better matching accuracy or robustness to noise can be achieved with higher order graph matching [4, 15, 24]. A common 3rd-order constraint expresses triangle similarity [15]. 4th-order constraints typically include consistency w.r.t. a local affine transformation [9, 28]. Graph matchers supporting edges of even higher order can for instance also express projective-invariant potentials [9]. However, despite recent advances in high-order graph matching, the running time and memory consumption remain an issue, especially for large datasets (images with hundreds or thousands of features): the complexity is at least $O(N^d)$ where N is the number of points and d the order. Besides, although some methods explicitly include a treatment of outliers, e.g., using absorbing nodes [15], the inlier rate is still assumed to be relatively large. For instance, Lee et al. [15] only describe experiments with at least 50% inliers (and at most 60 points), and Duchenne et al. [9] show a severe drop of performance when the inlier rate falls below 30% (with less than 100 points).

Our work. Feature descriptors provide 1st-order photometric information to estimate correspondence likelihood and identify potential matches. All other information used for matching is generally restricted to geometric information, i.e., relative point location in the image. This is the case for RANSAC methods and for most existing graph matchers. Although some authors mention possible extensions of graph matching potentials to photometric information [9], such uses are scarce and tend to translate into quasi-dense matching [10]. We propose here a novel, simple and efficient, 2nd-order photometric criterion.

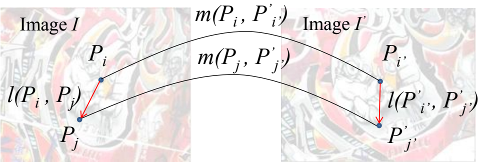

Figure 1: Lines (P_i,P_j) and $(P'_{i'},P'_{j'})$ are unlikely to be similar unless both matches $(P_i,P'_{i'})$ and $(P_j,P'_{j'})$ are correct.

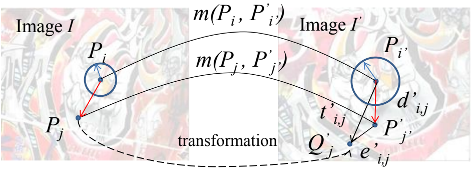

Figure 2: Distances used for the computation of the transformation error $\eta_{i,i',j,j'}$.

It is based on the fact that for points P_i,P_j in image I and $P'_{i'},P'_{j'}$ in image I', it is unlikely to find similar photometric information around lines (P_i,P_j) and $(P'_{i'},P'_{j'})$ unless both $(P_i,P'_{i'})$ and $(P_j,P'_{j'})$ are correct matches (see Fig. 1). To express this property, we define a virtual line descriptor (VLD) that captures photometric information between two points. The distance between two such descriptors measures the dissimilarity between the corresponding two virtual lines. It can be used in the 2nd-order term of graph matchers to improve their accuracy.

We also define a semi-local matching strategy based on VLD that considerably increases the inlier rate. It can be used as a preprocessor to RANSAC methods to improve the quality of match selection. As it can eliminate false matches near epipolar lines, it greatly improves precision. It also considerably reduces the number of iterations, which improves speed.

2 Virtual line descriptor (VLD)

The general idea of our descriptor for a virtual line between points P_i and P_j is to consider a regular covering, with some overlap, of an image strip between P_i and P_j, and use a SIFT-like descriptor to characterize each element of the covering. The global line descriptor is basically the concatenation of the descriptors of each covering element. It inherits SIFT descriptor's robustness to noise and changes of scale, orientation and illumination.

Geometric consistency. Before describing a line, we actually first check a geometric constraint, extending that of Albarelli et al. [1]. Given matches $m_{i,i'} = (P_i,P'_i)$ and $m_{j,j'} = (P_j,P'_{j'})$, and assuming that the local transformation around P'_i is close to a similitude, we define the point Q'_j in image I' as the expected position of $P'_{j'}$ (cf. Fig. 2):

$$Q'_j = P'_{i'} + \frac{s(P'_{i'})}{s(P_i)} R(a(P'_{i'}) - a(P_i)) \overrightarrow{P_iP_j} \qquad (1)$$

where $s(P)$ is the scale of feature point P, $a(P)$ is the main orientation at P, and $R(\alpha)$ is the rotation of angle α. Permuting I and I' defines a point $Q_{j'}$ in image I as the expected position of P_j. The transformation error of $(P_j,P'_{j'})$ by $(P_i,P'_{i'})$ is measured based on distances $d_{i,j} = dist(P_i,P_j)$, $t_{i,j} = dist(P_i,Q_j)$, $e_{i,j} = dist(P_j,Q_j)$, and likewise in I'. The normalized and symmetrized score of geometric consistency for matches $m_{i,i'}$ and $m_{j,j'}$ is defined as:

$$\chi(m_{i,i'},m_{j,j'}) = \min(\eta_{i,i',j,j'},\eta_{j,j',i,i'}) \quad \text{where} \quad \eta_{i,i',j,j'} = \frac{e_{i,j}}{\min(d_{i,j},t_{i,j})} = \frac{e'_{i',j'}}{\min(d'_{i',j'},t'_{i',j'})} \qquad (2)$$

Matches $m_{i,i'}$ and $m_{j,j'}$ are considered as *consistent w.r.t. geometry* iff $\chi(m_{i,i'},m_{j,j'}) < \chi_{\max}$. In all our experiments, we use a threshold value $\chi_{\max} = 0.5$. This fast prefiltering step eliminates many false matches before photometric comparison while preserving most inliners.

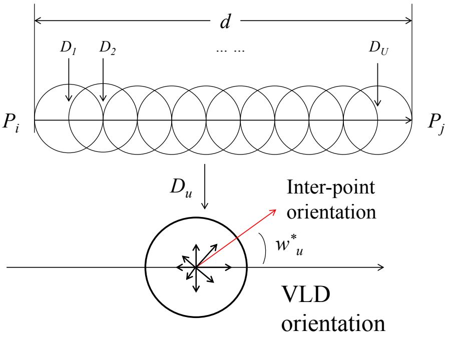

Figure 3: Top: disk covering of line (P_i, P_j). Bottom: 8-bin histogram of gradient orientation for disk D_u, and main orientation w^*_u.

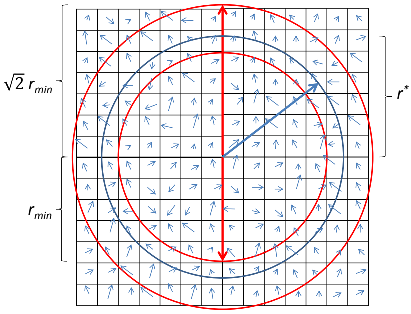

Figure 4: For $r \geq r_{\min}$, the VLD is computed on the q^{th} image scale on r^*-radius disks, where $r_{\min} \leq r^* = r/2^{q/2} < r_{\min}\sqrt{2}$.

Line covering. For any two points P_i and P_j in image I at distance d one from another, we consider U inter-point disks (D_u) of radius $r = \frac{d}{U+1}$ centered on points $P_i + \frac{u}{U+1}\overrightarrow{P_iP_j}$ for $u \in \{1,\ldots,U\}$ (see Fig. 3). Each disk is then described at image scale $s = \max(r/r_{\min}, 1)$ where $r_{\min}$ is a minimum description radius. In our experiments, we use $U = 10$ and $r_{\min} = 5$ pixels, which provides a good balance between discrimination and repeatability. In practice, scales can be discretized and precomputed, to avoid rescaling the image for each new pair of points. As for SIFT [18], we construct a pyramid of scaled images. In our experiments, a geometric progression with ratio $\sqrt{2}$ proved enough for repeatability. For any disk radius r in the original image I, we thus use scale $s^* = 2^{q/2}$ for q natural integer such that $2^{q/2} \leq s < 2^{(q+1)/2}$, i.e., $q = \lfloor 2\frac{\log s}{\log 2}\rfloor$. In the scaled image, the disk radius is $r^* = r/s^*$ (see Fig. 4).

Inter-point gradient histogram. The descriptor for disk D_u is a single SIFT-like local gradient histogram [18]. (SIFT actually defines a grid of 4×4 such histograms.) We use V bins $(h_{u,v})_{v\in\{1,\ldots,V\}}$ to represent the distribution: each pixel in the disk votes in the orientation bin corresponding to its gradient (relatively to the line direction), weighted by the gradient magnitude and by a Gaussian-weighted circular window with $\sigma = \frac{3}{2}r^*$. The line histograms are then normalized so that $\sum_{u=1}^{U}\sum_{v=1}^{V} h_{u,v} = 1$. In our experiments, like SIFT, we use $V = 8$.

Inter-point orientation. Inter-point orientation is computed as SIFT too, with some adaptation. We construct an orientation histogram for D_u using W bins $(O_{u,w})_{w\in\{0,\ldots,W-1\}}$. As slightly more variations can be expected on the line between two points than on the feature point themselves, we recommend $W > V$. In all our experiments we use $W = 24$ (as opposed to 36 for SIFT). In addition, we treat opposite direction together and actually consider the derived histogram $(\widehat{O}_{u,w})_{w\in\{0,\ldots,W-1\}}$ defined as $\widehat{O}_{u,w} = O_{u,w} - O_{u,(w+W/2)\,\mathrm{mod}\,W}$. (Only $W/2$ bins are actually needed to accumulate gradient information.) This reduction of the number of bins preserves enough discrimination while enhancing robustness. The main orientation w^*_u is finally defined as the bin of the derived histogram with the highest value. We also define normalizing factors γ_u to weight each orientation over the whole line:

$$w^*_u = \underset{w\in\{0,\ldots,W-1\}}{\operatorname{argmax}} \widehat{O}_{u,w} \qquad \gamma_u = \frac{\widehat{O}_{u,w^*_u}}{\sum_{u=1}^{U}\widehat{O}_{u,w^*_u}} \geq 0 \tag{3}$$

Each disk D_u is represented by the histogram $(h_{u,v})_{v\in\{1,\ldots,V\}}$, the orientation w^*_u and the normalizing factor γ_u. The size of the overall line descriptor is thus $U\,(V+2)$.

High contrast suppression. Lines with high contrast might be unreliable, in particular along image edges. In this case, gradients have strong responses in the orthogonal direction, which leads VLDs to be similar and thus not discriminative. For this reason, we define a line contrast indicator: $\kappa = \frac{s^*}{U\,d} \sum_{u=1}^{U} \widehat{O}_{u,w_u^*}$. VLDs with contrast κ above given threshold $\kappa_{\max}$ are considered unreliable and discarded. Experimentally, assuming image intensity in the range 0..255, we use $\kappa_{\max} = 30$. (Low contrast lines can also be discarded but are very unlikely.)

Distance between two VLDs. Given lines $l_{i,j} = (P_i, P_j)$ in I and $l'_{i',j'} = (P_{i'}, P_{j'})$ in I', we now define the distance between their descriptors. For each inter-point disk D_u in I and corresponding inter-point disk D'_u in I', we compute both the difference of the gradient histograms and the difference of the main orientations. Orientations w_u^* and $w_u'^*$ are compared modulo $W/2$ (most dissimilar orientation), normalized to 1, and weighted by the average of the orientation normalizing factors γ_u and γ'_u, resulting in a value in $[0,1]$. The differences of gradient histograms and main orientations are then linearly combined with factor $\beta \in [0,1]$:

$$\tau(l,l') = \beta \sum_{u=1}^{U} \sum_{v=1}^{V} |h_{u,v} - h'_{u,v}| + (1-\beta) \sum_{u=1}^{U} \left(\frac{\gamma_u + \gamma'_u}{2} . \frac{\min(|w_u^* - w_u'^*|, W - |w_u^* - w_u'^*|)}{W/2} \right) \qquad (4)$$

VLD-consistency. The VLD-distance between matches $m_{i,i'} = (P_i, P_{i'})$ and $m_{j,j'} = (P_j, P_{j'})$ is the VLD-distance between the corresponding lines: $\tau(m_{i,i'}, m_{j,j'}) = \tau(l_{i,j}, l_{i',j'}) \in [0,1]$. It can be used in the pairwise score of a graph matcher, e.g., with a contribution of the form $\exp(-\lambda\tau^2)$. Experimentally, we use $\beta = 0.36$ and $\lambda = 100$. When an all-or-nothing choice is required, matches $m_{i,i'}$ and $m_{j,j'}$ are said *VLD-consistent* iff their virtual lines are valid w.r.t. contrast and $\tau(m_{i,i'}, m_{j,j'}) \leq \tau_{\max}$. In our experiments, we use $\tau_{\max} = 0.35$. The matches are said *gVLD-consistent* iff they are both geometry- and VLD-consistent.

Parameters. Our descriptor may seem to have many parameters but most of them actually are directly imported from SIFT. Experiments just taught us that SIFT standard values could be weakened here for a lighter but still discriminant and robust descriptor. The only specific VLD parameters are the number of disks (U, with minimum radius $r_{\min}$), the balance between gradients and orientations (β) and a possible weight when used for a pairwise score (λ). Parameters such as β and λ could be learned as in [17]. K-VLD below uses additionally consistency thresholds for geometry ($\chi_{\max}$), contrast ($\kappa_{\max}$) and photometry ($\tau_{\max}$).

3 K-VLD: a K-connected VLD-based matching method

VLD can be directly used as a pairwise constraint in 2nd or higher-order graph matching methods. Yet, existing graph matching methods do not scale well to large numbers of matches and, as shown in the experiment section, may perform poorly when the foreground creates background occlusions. Besides, some of them are not well suited for large outlier elimination. In this section, we introduce K-VLD, a novel matching method that overcomes these limitations. It is semi-local in the sense that the score of a match depends on its consistency with neighboring matches. The consistency is both geometric and photometric, using our VLD criterion. The basic idea is that, given a potential match $(P_i, P'_{i'})$, if there are in the neighborhood of P_i and $P'_{i'}$ at least K other matches $(P_{j_k}, P'_{j'_k})_{k \in \{1,\ldots,K\}}$ that are gVLD-consistent with $(P_i, P'_{i'})$, then $(P_i, P'_{i'})$ is likely to be a correct match. The method can be seen as a simplified 2nd-order graph matcher specialized for image features. It provides a binary assessment for each match (correct or not) as well as a consistency score for further filtering.

Neighborhoods. K-connected gVLD-consistency has to apply only within a neighborhood of the points. We adapt the size of the neighborhoods to the density ρ of feature points. Neighborhoods are defined as disks centered on P_i and $P'_{i'}$, with respective radius B and B'. Given a set $\mathcal{M}$ of potential matches between I and I', with minimum inlier rate $\rho_{\min}$, then the average number of correct matches K_B in a B-neighborhood is:

$$K_B = \frac{\pi B^2}{\text{area}(I)} \rho_{\min} |\mathcal{M}| \qquad B_K = \sqrt{\frac{K\,\text{area}(I)}{\pi\, \rho_{min}\, |\mathcal{M}|}} \tag{5}$$

As B should be chosen such that $K_B \geq K$, we get a definition for the minimum radius B_K of the neighborhood (see Eq. 5). Moreover, for stability reason, we exclude neighboring points P_j that are too close to P_i, within $B_{\min}$ pixels. Wrapping up, we say that a match $(P_j, P'_{j'})$ is a neighbor of $(P_i, P'_{i'})$ iff P_j is in the $(B_{\min}, B_K)$-annulus centered on P_i in I, or $P'_{j'}$ is in the $(B_{\min}, B'_K)$-annulus centered on $P'_{i'}$ in I'. (The definition is symmetrical.) In the unlikely case that we discover a posteriori that $\rho < \rho_{\min}$, B_K has to be expanded accordingly and the algorithm has to be rerun. In all our experiments, we set $\rho_{\min} = 3\%$ and $B_{\min} = 10$ pixels. Then, given a set of matches M and a match $m \in M$, we define the following neighborhoods:

- $\mathcal{N}_M(m) = \{m' \in M \mid m \text{ and } m' \text{are neighbors}\}$
- $\mathcal{N}_{M,\text{geom}}(m) = \{m' \in M \mid m \text{ and } m' \text{are geometry-consistent neighbors}\} \subset \mathcal{N}_M(m)$
- $\mathcal{N}_{M,\text{gvld}}(m) = \{m' \in M \mid m \text{ and } m' \text{are gVLD-consistent neighbors}\} \subset \mathcal{N}_{M,\text{geom}}(m)$

Problem statement. Experimentally, requiring that good matches have at least K gVLD-consistent neighbors eliminates many outliers, but some may still remain, especially in ambiguous scenes. We found that adding an extra constraint on the proportion of geometry-consistent neighbors helped in removing many of these remaining outliers. More formally, given a set of potential matches $\mathcal{M}$, we look for a subset $M \subset \mathcal{M}$ such that, for all $m \in M$,

$$|\mathcal{N}_{M,\text{gvld}}(m)| \geq K \quad \text{and} \quad \left(\frac{|\mathcal{N}_{M,\text{geom}}(m)|}{|\mathcal{N}_M(m)|} \geq \omega_{\min} \quad \text{or} \quad \frac{\sum_{m' \in \mathcal{N}_M(m)} \chi(m,m')}{|\mathcal{N}_M(m)|} \leq \bar{\chi}_{\max} \right) \tag{6}$$

We are actually interested in a set M^* with maximum cardinality satisfying this condition. The absence of ambiguous matches in M can also be imposed (see below).

Algorithm. For efficiency reasons, we actually only look for sets M of large cardinality satisfying equation (6). Our algorithm starts with $M = \mathcal{M}$ and repeatedly performs:

1. remove all $m \in M$ such that $N_{M,\text{gvld}}(m) < K$
2. remove all $m \in M$ such that $\frac{|\mathcal{N}_{M,\text{geom}}(m)|}{|\mathcal{N}_M(m)|} < \omega_{\min}$ and $\frac{\sum_{m' \in \mathcal{N}_M(m)} \chi(m,m')}{|\mathcal{N}_M(m)|} > \bar{\chi}_{\max}$

until no match is removed. Upon termination, which always occurs as $|M|$ strictly decreases (after around 3 iterations, at most 5 in practice), either $M = \emptyset$ or M satisfies condition (6). In practice, this yields a large set of matches for M, almost never empty. gVLD-consistency is enforced first because it is the strongest condition; the value of M is thus best estimated when performing step 2. In all our experiments, we use $K = 3$, $\omega_{\min} = 30\%$ and $\bar{\chi}_{\max} = 1.2$.

Dealing with ambiguity. Ambiguous matches, i.e., matches that share a point, correspond to a special kind of outliers. Treating them as ordinary matches when eliminating outliers does not guarantee an ambiguity-free set of final matches. For this reason, a heuristic elimination is often performed to keep at most one match per point, generally by keeping only the one with the best score. But sometimes there is no easy choice, e.g., with ambiguous points on an epipolar line. In that case we can use VLD information to improve disambiguation. For this, we sort the matches in M so that matches m with highest number of gVLD-consistent neighbors are preferred or, if equal, highest average of VLD score among these gVLD-consistent neighbors, i.e., $T(m) = \sum_{m' \in \mathcal{N}_{M,\mathsf{gvld}}} \tau(m,m')/|\mathcal{N}_{M,\mathsf{gvld}}(m)|$. More formally, matches are sorted (less likely matches first) according to the order relation: $m \preccurlyeq_M m'$ iff

$$|\mathcal{N}_{M,\mathsf{gvld}}(m)| < |\mathcal{N}_{M,\mathsf{gvld}}(m')| \;\text{ or }\; \big(|\mathcal{N}_{M,\mathsf{gvld}}(m)| = |\mathcal{N}_{M,\mathsf{gvld}}(m')| \;\text{ and }\; T(m) \geq T(m')\big) \quad (7)$$

And we add a third step to the algorithm:

3. for each $m \in M$ in $\preccurlyeq_M$-sort order, remove m if $\exists m' \in M \backslash \{m\}$ s.t. m' conflicts with m

It is to be performed only once as all conflicts get removed and none can later be created. The sorting can also be used in step 1 and 2 of the algorithm for picking $m \in M$, updating M as matches are removed. However, it does lead to much better results experimentally.

Optimizations and heuristics. Given a match m, we need to count the number of gVLD-consistent neighbors m', which requires computing $\tau(m,m')$. To avoid recomputation, we keep theses values in a cache. Besides, to speedup the algorithm, we do not have to enumerate all gVLD-consistent neighbors. It is enough to stop after $N_{\max} \gg K$ neighbors are found, as m is then extremely unlikely to be later removed (e.g., $N_{\max} = 20$ for $K = 3$).

4 Evaluation

We experimented with existing matching methods, we augmented some of them with VLD, and we compared with K-VLD. We evaluated matching accuracy in various imaging conditions. We also tested K-VLD as a prefilter to RANSAC-based calibration.

All extracted features are SIFT keypoints, as implemented in VLFeat [27]. We selected a range of state-of-art methods presenting the rich variety of graph matching methods. We use the authors' code for probabilistic hypergraph matching (HGM) [28], hypergraph matching via reweighted random walks (RRWHM) [15] and tensor matching (TM) [9], and our own implementation of spectral matching (SM) [16], which computes the same matching results as integer projected fixed point (IPFP) [17], and game-theoretic matching (GTM) [1]. VLD is incorporated to SM (2^{nd}-order method) and HGM. For calibration, we used the IPOL implementation of ORSA [22, 23], which is a parameterless, state-of-art RANSAC variant.

Changing imaging conditions. We use Mikolajczyk's dataset, that evaluates feature detectors and descriptors under different image transformations, including change of viewpoint and illumination, zoom, blur and rotation [21]. It is composed of 8 sequences of 6 images with increasing variation. For each sequence, we successively match image 1 with the others. We extract the best 400 SIFT matches (i.e., with lowest descriptor distance) as candidates for each image pair. 400 features was about the limit that methods TM and RRWHM could handle on a 24 GB computer, running in 200 s; K-VLD runs in 1s, with a performance quasi linear in $|\mathcal{M}|$. For each method, we extract the N best matches according to the method, where N is the number of ground truth inliers. Matches with less than 5-pixel transformation error are considered as inliers, and accuracy is the proportion of inliers among the N returned matches [15]. This dataset features image transformations that can be described by a single homography; as lines are preserved, VLDs are expected to be relatively stable. In fact, results in Fig. 5 show that K-VLD often outperforms other methods. Besides, VLD significantly improves existing methods, especially for scenes with viewpoint or scale changes.

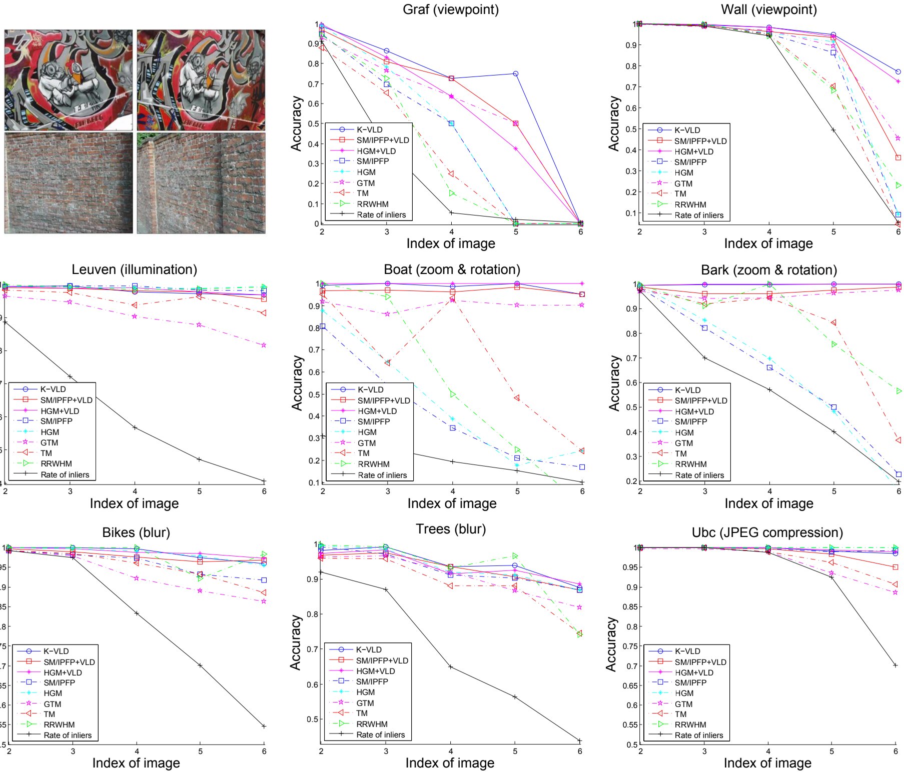

Figure 5: Matching accuracy measured on Mikolajczyk's dataset.

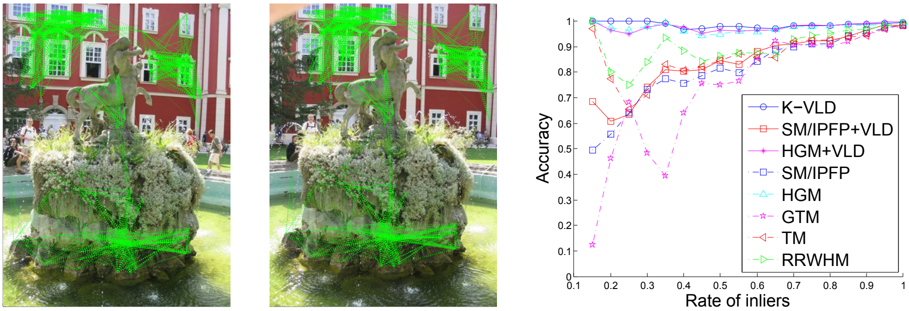

Figure 6: Dětenice fountain: K-VLD clusters (1 image pair) & average accuracy on all pairs.

Strong occlusions. To evaluate the case of occlusions, we use the Dětenice fountain dataset [7], from which we took a sequence of 43 images. The occluding foreground (a statue) creates strong variations in the background. A ground truth calibration is first constructed by selecting 50 inliers by hand. As above, we extract the best 400 SIFT matches for each pair of successive images. We then measure the actual inlier rate and the accuracy. Results are shown in Fig. 6, where strong local variations have been smoothed and resampled in plotted graphs for readability. K-VLD creates clusters of consistent matches despite occlusions, outperforming other methods most of the time. VLD improves SM moderately (5–10% more inliers) and HGM only slightly, as it already has an excellent performance.

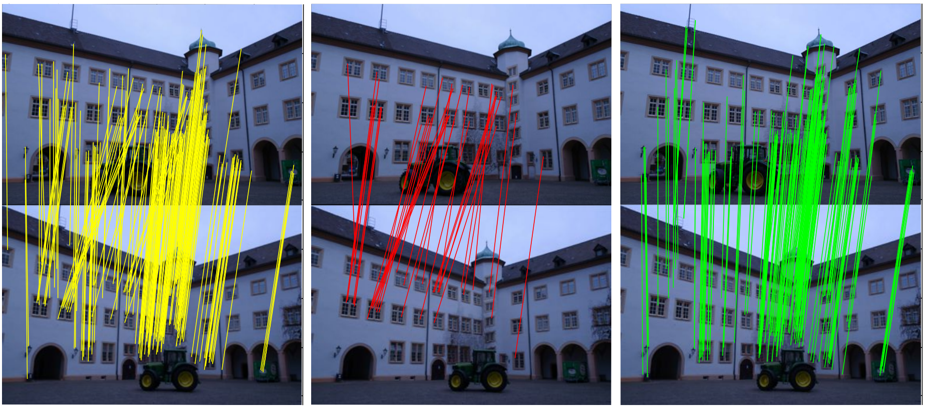

Figure 7: Left: inliers by ORSA. Middle: false matches near epipolar lines not eliminated by ORSA but rejected by K-VLD. Right: inliers by K-VLD + ORSA. (Symmetric matches; Lowe criterion threshold = 0.8; only 1/4 matches shown, thus matches may show or hide.)

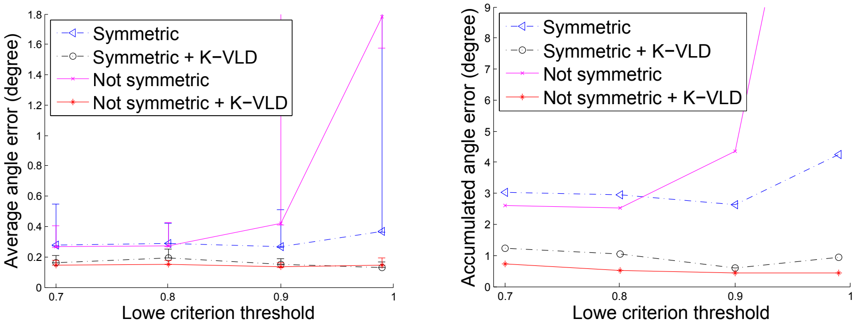

Figure 8: Angle error on the castle dataset. Left: average error over 19 image pairs. Error bars indicate standard deviation (not min/max). Right: accumulated error after one loop.

Ambiguity and RANSAC prefiltering. To evaluate the benefits of K-VLD as a preprocessing filter for RANSAC-based calibration, we use Strecha's castle dataset [25]. It is composed of a looping sequence of 19 images of a courtyard and provides ground truth for both internal and external camera calibration. The scene is highly ambiguous due to many repeated windows, which degrades registration, as illustrated on Fig. 7. Around 5,000 to 7,000 SIFT points are extracted in each image. Potential matches are generated for each pair of consecutive images and for different values of the Lowe rejection threshold [18], i.e., the distance ratio of closest to second closest keypoint (often 0.8 in the literature). We also test the case of symmetric matches only, i.e., points P whose match P' in the second image has P as match in the first image. This yields 300 to 3,000 matches per pair. We measure the average angle error over each pair and the error standard deviation. As the last image can be compared with the first one, we also measure the accumulated angle error independently of the ground truth by multiplying all the rotation matrices and measuring the angle difference with identity. We compare two methods for estimating the fundamental matrix: using ORSA alone, or prefiltering the matches with K-VLD before ORSA. Results are shown in Fig. 8. The use of K-VLD as a match prefilter greatly improves stability (deviation) and precision.

5 Conclusion

For 2nd-order graph matching, distinctive pairwise constraints are crucial, just as distinctiveness is crucial for ordinary, 1st-order feature matching. As our experiments show, our virtual line descriptor VLD provides such distinctiveness, offering a better accuracy to 2nd-order graph matchers and preserving the scalability (time and space) to a large number of points contrary to high-order graph matchers. Besides, compared to other graph matching methods, our K-VLD matcher provides among the best accuracy, especially when the inlier rate drops, even down to a few percents (less than 5 or 10%) and despite strong occlusions (thanks to its semi-local nature) or strong viewpoint or scale changes. Moreover, used as a preprocessor to RANSAC, K-VLD eliminates most outliers, including near epipolar lines and despite possible ambiguities, which greatly improves calibration precision.

These results were achieved with unchanged parameter values after experimenting a few options, but a more systematic study *à la* SIFT [18] on a larger benchmark would be valuable. Interestingly, K-VLD only uses 2nd-order constraints; 1st-order terms could also be added.

References

[1] A. Albarelli, E. Rodola, and A. Torsello. Robust game-theoretic inlier selection for bundle adjustment. In *3DPVT*, 2010.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[3] Alexander C. Berg, Tamara L. Berg, and Jitendra Malik. Shape matching and object recognition using low distortion correspondence. In *CVPR*, pages 26–33. IEEE, 2005.

[4] M. Chertok and Y. Keller. Efficient high order matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(12):2205–2215, 2010.

[5] Minsu Cho, Jungmin Lee, and Kyoung Mu Lee. Reweighted random walks for graph matching. In *European Conference on Computer Vision (ECCV)*, pages 492–505, 2010.

[6] Ondrej Chum and Jiri Matas. Matching with PROSAC – progressive sample consensus. In *Conf. Computer Vision & Pattern Recognition (CVPR)*, pages 220–226. IEEE, 2005.

[7] CMP, Czech Technical University, Prague. Dětenice fountain dataset, 2008. http://cmp.felk.cvut.cz/projects/is3d/Data.html.

[8] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *IJPRAI*, 18:265–298, 2004.

[9] Olivier Duchenne, Francis Bach, In-So Kweon, and Jean Ponce. A tensor-based algorithm for high-order graph matching. *PAMI*, 33(12):2383–2395, December 2011.

[10] V. Ferrari, T. Tuytelaars, and L. Van Gool. Simultaneous object recognition and segmentation by image exploration. *ECCV*, pages 40–54, 2004.

[11] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[12] P. E. Forssén and D. G. Lowe. Shape descriptors for maximally stable extremal regions. In *11th IEEE Int'l Conference on Computer Vision (ICCV)*, pages 1–8. IEEE, 2007.

[13] S. Gu, Y. Zheng, and C. Tomasi. Critical nets and beta-stable features for image matching. In *European Conference on Computer Vision (ECCV)*, pages 663–676, 2010.

[14] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. ISBN 0521540518.

[15] J. Lee, M. Cho, and K.M. Lee. Hyper-graph matching via reweighted random walks. In *IEEE Conf. Computer Vision & Pattern Recognition (CVPR)*, pages 1633–1640, 2011.

[16] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *10th ICCV*, volume 2, pages 1482–1489. IEEE, 2005.

[17] M. Leordeanu, R. Sukthankar, and M. Hebert. Unsupervised learning for graph matching. *International Journal of Computer Vision (IJCV)*, 96(1):28–45, 2012.

[18] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

[19] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[20] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International Journal of Computer Vision (IJCV)*, 60(1):63–86, 2004.

[21] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, 27(10):1615–1630, October 2005.

[22] L. Moisan and B. Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *IJCV*, 57(3):201–218, 2004.

[23] L. Moisan, P. Moulon, and P. Monasse. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line (IPOL)*, 2012. http://www.ipol.im/pub/algo/mmm_orsa_homography.

[24] P. Ren, R. C. Wilson, and E. R. Hancock. High order structural matching using dominant cluster analysis. In *Image Analysis and Processing (ICIAP)*, 2011.

[25] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[26] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[27] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. In *Int'l Conference on Multimedia*, pages 1469–1472. ACM, 2010.

[28] R. Zass and A. Shashua. Probabilistic graph and hypergraph matching. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.