Efficient and Scalable 4th-order Match Propagation

David Ok, Renaud Marlet, and Jean-Yves Audibert

Université Paris-Est, LIGM (UMR CNRS), Center for Visual Computing
École des Ponts ParisTech, 6-8 av. Blaise Pascal, 77455 Marne-la-Vallée, France
Email: {david.ok@imagine.enpc.fr, renaud.marlet@enpc.fr, jyaudibert@gmail.com}

Outline

In this supplementary material, the following points are elaborated and additional experimental data are provided.

1. Predicate $\mathbb{P}_f$ for position, shape, and orientation consistency (§2.1 of paper).
2. Distrust score (§2.3 of paper).
3. Proof of properties regarding region affine-consistency (§2.3 of paper).
4. Deformable Object Matching (§4.1 of paper).
5. Accurate and Scalable Matching for Camera Calibration (§4.2 of paper).
6. Accurate Pattern Localization: Window Detection (§4.3 of paper).
7. Affinity Estimation: Triple vs Single Match (§4.4 of paper).

1 Predicate $\mathbb{P}_f$ for Position, Shape, and Orientation Consistency

All our reported experiments involve feature points with known scales (DOG, Harris-Affine, Hessian-Affine and MSER, all described with SIFT). In this case, position consistency is somehow already taken into account within the shape consistency condition. We thus did not include condition $d_{x_1}(\mathbf{x}_2) < \delta_\mathrm{p}$ in $\mathbb{P}_f$ (see condition (2) in paper, line 158).

Regarding shape and orientation consistency, we used $\delta_\mathrm{s} = 0.4$ and $\delta_\mathrm{o} = 60°$ for all four kinds of features. In fact, such permissive thresholds are required for DoG+SIFT, whose shape and orientation is not very precise, but they could actually have been stricter for the other features, especially regarding orientation.

2 Distrust Score

The distrust score is illustrated in fig. 1.

With this score, ambiguous matches tend to be ordered after unambiguous matches. Moreover, the search may be efficiently pruned by putting an upper bound on distrust. Ambiguous matches with larger distrust are excluded right from the start and are never considered for seeding or growing a region. They thus do not appear in the final selection of matches $\mathcal{R}$.

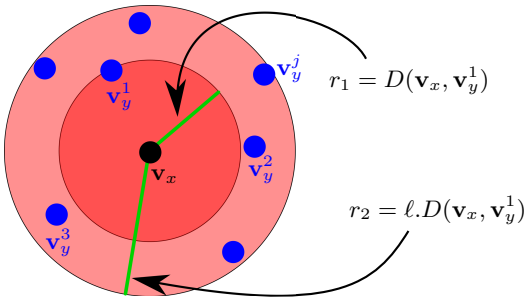


Fig. 1: Illustration of the distrust score L. Nearest neighbors $\{\mathbf{v}_y^i\}_{i=2}^j$ in $\mathcal{Y}$ to $\mathbf{v}_x$ lie in the annulus with center $\mathbf{v}_x$, inner radius $r_1 = D(\mathbf{v}_x, \mathbf{v}_y^1)$ and outer radius $r_2 = \ell.D(\mathbf{v}_x, \mathbf{v}_y^1)$. As a reminder: $L(x, y_1) \leq 1$ and $1 \leq L(x, y_i) \leq \ell$ for $i \geq 2$.

The resulting match ordering and filtering is used to always select the best match candidates first, either for constructing a region seed, i.e., a match quadruple, or for growing a region. (Also, because of the greedy strategy, we only consider matches that are not currently assigned to a region.)

3 Proof of Properties Regarding Region Affine-Consistency

We sketch the proofs of the two properties expressed in the paper regarding the affine-consistency of the union of affine-consistent regions (§ 2.3, *Region Affine-Consistency*, line 216) and the unicity of minimal subpartition into affine-consistent regions with maximum size (§ 2.3, *Maximal Consistency*, line 224).

Affine-Consistency of the Union of Affine-Consistent Regions (line 216). Given two affine-consistent regions R, R' such that $R \cap R' \neq \varnothing$, then $R \cup R'$ is affine-consistent. To show this, consider two matches $m, m' \in R \cup R'$. If m and m' are both in R (resp. both in R'), then m, m' are affine-consistent in R (resp. in R'), hence in $R \cup R'$. If $m \in R$ and $m' \in R'$, let $m_0 \in R \cap R'$. Then there exists a sequence of affine-consistent quadruples in R from m to m_0 and a sequence of affine-consistent quadruples in R' from m_0 to m'. The concatenation of both sequences form a sequence of affine-consistent quadruples in $R \cup R'$ from m to m'. ■

Maximal Consistency (line 224). If a subpartition of a set of matches $\mathcal{M}$ into affine-consistent regions with maximum size is such that the number of regions is minimal, then the subpartition is also unique (i.e., the only one with this property). To show this, consider $\mathcal{R}, \mathcal{R}'$ two subpartitions of $\mathcal{M}$ into affine-consistent regions such that $\|\mathcal{R}\| = \|\mathcal{R}'\|$ is maximal and $|\mathcal{R}| = |\mathcal{R}'|$ is minimal. First, there does not exist $R \in \mathcal{R}$ such that $R \cap (\bigcup_{R' \in \mathcal{R}'} R') = \varnothing$. Otherwise $\mathcal{R}'_0 = \mathcal{R}' \cup \{R\}$ would be a subpartition of $\mathcal{M}$ into affine-consistent regions such

that $\|\mathcal{R}'_0\| > \|\mathcal{R}'\|$, which contradicts the maximality assumption. Thus for any $R \in \mathcal{R}$, there exists $R' \in \mathcal{R}'$ such that $R \cap R' \neq \varnothing$. But then $R = R'$. Otherwise consider $m \in R' \setminus R$, form $R_0 = R' \cup \bigcup_{R_i \in \mathcal{R}, R_i \cap R' \neq \varnothing} R_i$, which is affine-consistent according to the above property (cf. region affine-consistency), and define $\mathcal{R}_0$ as $\mathcal{R}$ where all the R_i occurring for constructing R_0 (including R) are replaced by just R_0. Then $\mathcal{R}_0$ is a subpartition of $\mathcal{M}$ into affine-consistent regions, whose size is at least that of $\mathcal{R}$: $\|\mathcal{R}_0\| \geq \|\mathcal{R}\|$. Now if $m \in R' \setminus R$ is in some R_i, then $R_i \neq R$ and at least R and R_i are merged in R_0 to form $\mathcal{R}_0$, hence $|\mathcal{R}_0| < |\mathcal{R}|$ which contradicts the minimality assumption of the cardinality of regions. And if $m \in R' \setminus R$ is in no R_i, then $|R_0| > |\bigcup_{R_i \in \mathcal{R}, R_i \cap R' \neq \varnothing} R_i| = \sum_{R_i \in \mathcal{R}, R_i \cap R' \neq \varnothing} |R_i|$, hence $\|\mathcal{R}_0\| > \|\mathcal{R}\|$, which contradicts the size maximality assumption. Symmetrically, for any $R' \in \mathcal{R}'$, there exists $R \in \mathcal{R}$ such that $R = R'$. Thus $\mathcal{R} = \mathcal{R}'$. ■

4 Deformable Object Matching

The ETHZ Toys dataset can be found and downloaded from [1]. We show in fig. 2 the results that we obtained on the 23 test images, which contain 43 objects in total. The names in italic such as *All* refer to the image file names in the dataset. In the illustrations, we use different colors to reference each model object, as summarized in table 1. Our results are shown for Harris-Affine and MSER interest points. For each image pair, we obtained matches such that their distrust score is less than $\ell = 1.1$, whereas [2] uses a much more restricted set of matches, i.e., such that their distrust score is less than $\ell = 0.9$.

We can see that the results are visually very clean. The consistency of matches actually goes beyond the mere recognition of objects. As already pointed out in the paper, region affine-consistency does not assume a single affinity but many. And this piecewise affinity is flexible enough to adapt to substantial non-affine transformations, as occurs with deformable objects.

Color	Model object	Color	Model object
Red	Blonde	Green	Car
Blue	Guard	Magenta	Leo
Yellow	Michelle	Orange	Ovo
Dark Green	Suchard	Lavender Blue	Xmas

Table 1: Reference color for each model object.

5 Accurate and Scalable Matching for Camera Calibration

In fig. 3a and fig. 3b, we show the 3D point clouds corresponding to the datasets *Books* and *Mars* [3], as obtained by our matching method followed by Bundler.

6 Accurate Pattern Localization: Window Detection

Pattern detection has been extensively studied and we refer the reader to the exhaustive survey of Liu et al. [4] for details. In pattern detection, structural regularity or symmetry is commonly assumed [4]. However, it is not always appropriate, such as for the eTRIMS dataset [5], and this makes our method complementary to most of the methods surveyed in [4].

We experimented indeed our method with the eTRIMS dataset. In this experiment, we considered initial matches such that their distrust score is less than $\ell = 1.2$, i.e. matches within 20% of the best match (descriptor-wise). Excerpts of our results are shown in fig. 4 and 5. The robustness of our pattern detection in the non-rectified case shows that a preliminary rectification (often defined manually and to be repeated for each different facade plane in a given image) is not absolutely necessary.

7 Affinity Estimation: Triple vs Single Match

Mikolajczyk et al. describe in [6] how to compute affinites from single matches. In addition to precision rates already depicted in fig. 4 of paper, we show the corresponding recall rates in fig. 6 . As already mentioned in the paper, precision rates clearly benefits from affinities computed with match triples, as opposed to single matches. Here we show that it is not at the expense of recall rates, that are comparable in the two approaches.

References

1. CALVIN research group: ETHZ toys datasets (2004) http://www.vision.ee.ethz.ch/~calvin/datasets.html.
2. Cho, M., Lee, J., Lee, K.M.: Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: ICCV. (2009)
3. Pajdla, T., Heller, J., Kukelova, Z., Havlena, M.: PRoVisG Mars 3D Challenge (2011) http://cmp.felk.cvut.cz/mars/.
4. Liu, Y., Hel-Or, H., Kaplan, C.S., Gool, L.J.V.: Computational symmetry in computer vision and computer graphics. Foundations and Trends in Computer Graphics and Vision **5** (2010) 1–195
5. Korč, F., Förstner, W.: eTRIMS Image Database for interpreting images of man-made scenes. Technical Report TR-IGG-P-2009-01, University of Bonn (2009)
6. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision **60** (2004) 63–86

(a) *All* (b) *CarBooks* (c) *CarGlas* (d) *CarViewpoint*

(e) *CarXmasA* (f) *CarXmasB* (g) *Magazines* (h) *GuardOnBlonde*

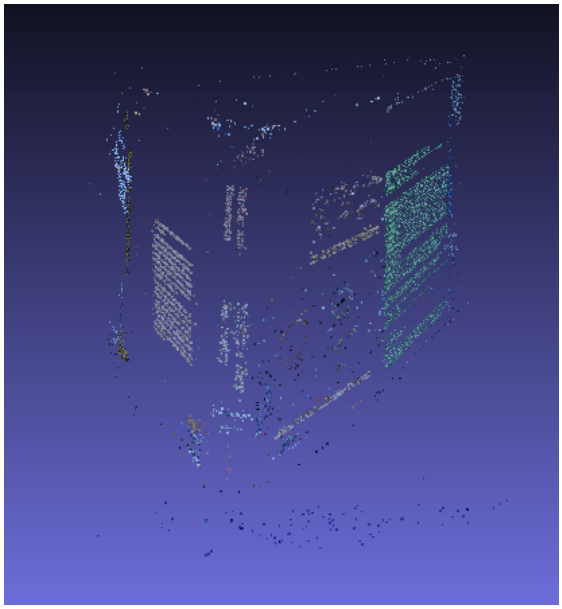
(i) *GuardOnMichelle* (j) *KellogsClutter* (k) *KitchenA* (l) *KitchenB*

(m) *LeoHidden* (n) *LeoSleeps* (o) *MichelleBentA* (p) *MichelleBentB*

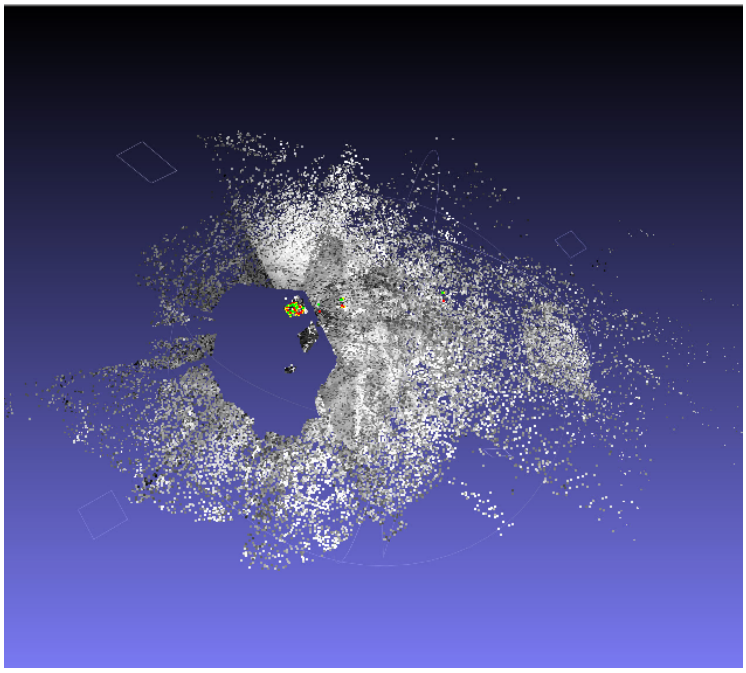
(q) *MichelleBentC* (r) *MichelleBentD* (s) *OvoLeoSuchardA* (t) *OvoLeoSuchardB*

(u) *OvoLeoSuchardC* (v) *TableA* (w) *TableB*

Fig. 2: Detection in ETHZ Toys test images. Colors are defined in Table 1.



(a) 3D point cloud for the *Books* dataset.



(b) 3D point cloud for the *Mars* dataset. The camera position appears in yellow, red and green.

Fig. 3: Reconstructed 3D point clouds.

Fig. 4: Window detection results on the eTRIMS dataset (1/2): input quadrilaterals, detections on rectified images, and detections on non-rectified images.

Fig. 5: Window detection results on the eTRIMS dataset (2/2): input quadrilaterals, detections on rectified images, and detections on non-rectified images.

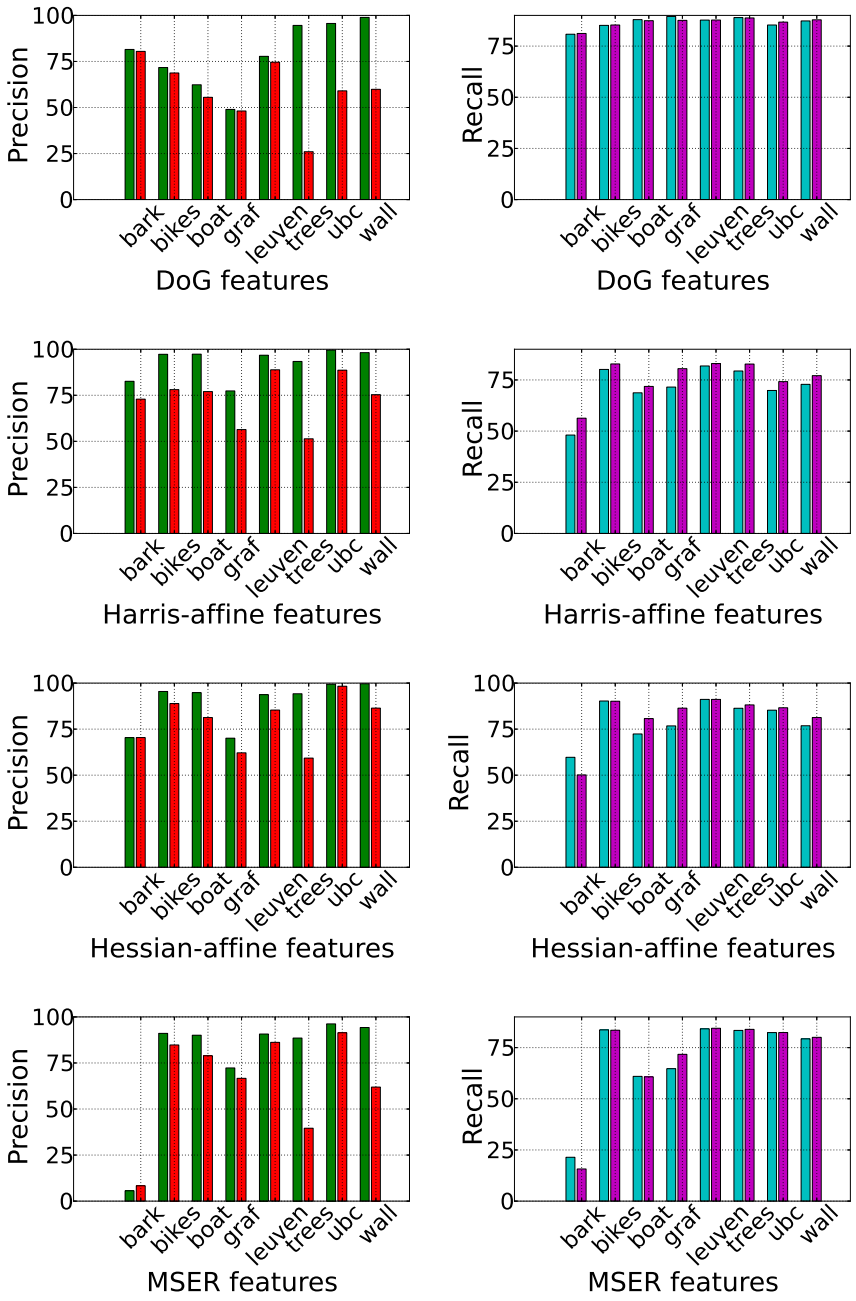


Fig. 6: Precision (%) and recall (%) of region growing on Mikolajczyk et al's dataset (pair 1-3).