Robust Spatial Matching as Ensemble of Weak Geometric Relations

Xiaomeng Wu
wu.xiaomeng@lab.ntt.co.jp

Kunio Kashino
kashino.kunio@lab.ntt.co.jp

NTT Corporation
3-1, Morinosato Wakamiya Atsugi-shi
Kanagawa, Japan 243-0198

Spatial matching methods including RANSAC, Hough transform and spatial context methods have been shown to be successful in addressing the mismatches resulted from the matching of local features in object retrieval. In these methods, true correspondences are identified by imposing a constraint on one or two classes of geometric coherences, e.g. in terms of spatial contexts or between-image transformations. These methods are potentially less discriminative due to the limited number of coherence classes, while forcibly enhancing the strength of constraints leads to the incorrect rejection of true correspondences. Spatial matching still faces a difficult trade-off between flexibility and discriminative power.

(a) Correspondences (b) Neighborhood coherence (c) Scaling coherence

(d) Rotation coherence (e) Normalization (f) Position coherence
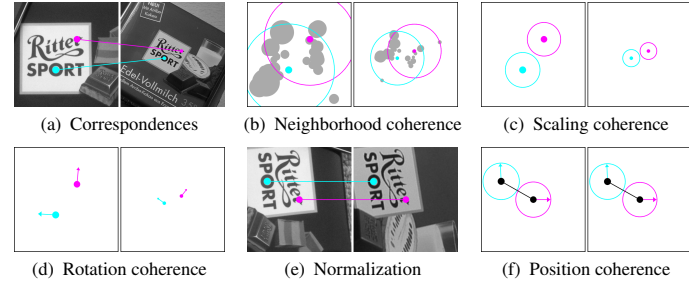
Figure 1: A correspondence pair and four geometric coherences.

To address this issue, we regard spatial matching as an ensemble of geometric relations on the set of feature correspondences. A *geometric relation* is a set of pairs of correspondences, in which every correspondence is associated with every other correspondence if and only if the pair satisfy a geometric constraint. We design a novel, unified collection of weak geometric relations that fall into four fundamental classes of geometric coherences (Figures 1(b)-1(d) and 1(f)) in terms of both spatial contexts and between-image transformations. By a *weak geometric relation*, we mean a sufficiently flexible constraint which, nevertheless, may offer only a limited discriminative power. The spatial similarity reduces to the cardinality of the conjunction of all geometric relations. The flexibility of weak geometric relations makes our method robust as regards rejections of true correspondences, and the conjunctive ensemble provides a high discriminative power in terms of mismatches.

An image is represented by a set P of features. For each feature $p \in P$ we are given its visual word $u(p)$, position $\mathbf{t}(p) = [x(p)\ y(p)]^T$, scale $\sigma(p)$ and orientation $\mathbf{R}(p)$. p can be mapped, from a unit circle heading a reference orientation, by a 3×3 transformation matrix $\mathbf{F}(p)$:

$$\mathbf{F}(p) = \begin{bmatrix} \mathbf{M}(p) & \mathbf{t}(p) \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (1)$$

where $\mathbf{M}(p) = \sigma(p)\mathbf{R}(p)$ is a linear transformation and homogeneous coordinates are to be used for the mapping. If $\sigma(p)$ is given by a real scalar, $\mathbf{F}(p)$ specifies a similarity transformation. $\mathbf{R}(p)$ is an orthogonal 2×2 matrix with $\det \mathbf{R}(p) = 1$, represented by an angle $\theta(p)$. Given two images P and Q, a correspondence $c \triangleq (p, q)$ is a pair of features $p \in P$ and $q \in Q$ with $u(p) = u(q)$. We assume $|C| \geq 2$ with $C = \{c\}$ and:

$$c = \left(u(c), \mathbf{t}(p), \sigma(p), \theta(p), \mathbf{t}(q), \sigma(q), \theta(q)\right). \quad (2)$$

Suppose that P and Q are related as regards a common near-rigid object and an unknown transformation $\mathbf{F}$. It can be inferred that all parts of the object obey the same transformation. Given a correspondence set C constructed from P and Q, there is a subset $C_\mathbf{F} \subseteq C$ of correspondences that lie inside the object and show similarity in terms of their local transformations. Spatial matching is to identify such a subset, whose cardinality provides evidence for the belief that P and Q include the same object.

We focus on the Cartesian product $C^2 = C \times C$, i.e. the set of all pairs (c_a, c_b) where $c_a, c_b \in C$. A function $h : C^2 \to \{0, 1\}$ is defined, which maps any arbitrary (c_a, c_b) to one if a geometric constraint is satisfied, and zero otherwise. A geometric relation G is thus a subset of C^2 such that $\forall (c_a, c_b) \in G$, $h(c_a, c_b) = 1$. If h is sufficiently well-defined and if the geometries in Eq. 2 are accurately given, we have $G \approx C_\mathbf{F}^2$. The spatial similarity can be formulated by the cardinality of G instead of that of $C_\mathbf{F}$.

Instead of a single h, we build a set $H = \{h\}$ of weak geometric constraints, resulting in a set $\mathcal{G} = \{G\}$ of geometric relations. Each $h \in H$ should be flexible as regards feature detection errors, but is allowed to offer a limited discriminative power. A conjunctive ensemble of such relations (Eq. 3) creates a single strong constraint that is expected to be discriminating in terms of mismatches. The spatial similarity thus becomes $|\hat{G}|$.

$$\hat{G} = \bigcap_{G \in \mathcal{G}} G = \left\{ (c_a, c_b) \in C^2 \,\middle|\, \left(\prod_{h \in H} h(c_a, c_b) \right) = 1 \right\} \quad (3)$$

More detail on our carefully designed weak geometric constraints is presented in the paper. Table 1 presents the reported MAPs of spatial matching methods on the OB, Paris and OB+F100K datasets. Our method outperforms all methods on all datasets. Our search time per query and per 1K images was 19.0 msec for OB. The shortest corresponding time so far reported by Shen et al. [4] was 17.6 msec. This reveals the high competitiveness of our scalability. Figure 2 compares HPM [2], the best baseline we know in this field, and our method.

Table 1: Comparison with reported, state-of-the-art MAPs.

Methods	OB	Paris	OB+F100K
Our Method	**.827**	**.766**	**.769**
HPM [2]	.789	.725	.730
Perdoch et al. [3]	.789	n/a	.726
Shen et al. [4]	.752	.741	.729
Arandjelovic and Zisserman [1]	.720	n/a	.642
Zhang et al. [5]	.713	n/a	.604
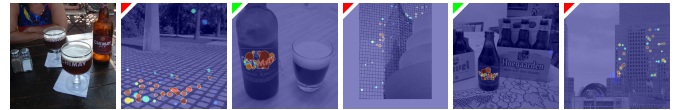
(a) Query (left) and top-five results returned by Hough pyramid matching [2].
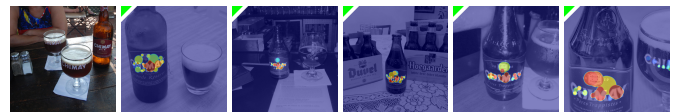
(b) Query (left) and top-five results returned by our method.

Figure 2: Comparison of HPM [2] and our method. The green and red colors of the upper-left corners of the images indicate positive and negative results. Identified correspondences are highlighted in colors.

[1] Relja Arandjelovic and Andrew Zisserman. Three things everyone should know to improve object retrieval. In *CVPR*, pages 2911–2918, 2012.

[2] Yannis S. Avrithis and Giorgos Tolias. Hough pyramid matching: Speeded-up geometry re-ranking for large scale image retrieval. *International Journal of Computer Vision*, 107(1):1–19, 2014.

[3] Michal Perdoch, Ondrej Chum, and Jiri Matas. Efficient representation of local geometry for large scale object retrieval. In *CVPR*, pages 9–16, 2009.

[4] Xiaohui Shen, Zhe Lin, Jonathan Brandt, and Ying Wu. Spatially-constrained similarity measure for large-scale object retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(6):1229–1241, 2014.

[5] Yimeng Zhang, Zhaoyin Jia, and Tsuhan Chen. Image retrieval with geometry-preserving visual phrases. In *CVPR*, pages 809–816, 2011.