Pattern Recognition Letters 73 (2016) 26–32

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# A general framework for image feature matching without geometric constraints☆

Jonas Toft Arnfred, Stefan Winkler*

*Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign (UIUC), 183632 Singapore*

ARTICLE INFO

*Article history:*
Received 22 January 2015
Available online 14 January 2016

*Keywords:*
Feature matching
Ratio match
Mirror match
Self match

ABSTRACT

Computer vision applications that involve the matching of local image features frequently use *Ratio-Match* as introduced by Lowe and others, but is this really the optimal approach? We formalize the theoretical foundation of *Ratio-Match* and propose a general framework encompassing *Ratio-Match* and three other matching methods. Using this framework, we establish a theoretical performance ranking in terms of precision and recall, proving that all three methods consistently outperform or equal *Ratio-Match*. We confirm the theoretical results experimentally on over 3000 image pairs and show that matching precision can be increased by up to 20 percentage-points without further assumptions about the images we are using. These gains are achieved by making only a few key changes of the *Ratio-Match* algorithm that do not affect computation times.



## 1. Introduction

Matching image points is a crucial ingredient in almost all computer vision applications that deal with sparse local image features, such as image categorization [5], image stitching [6], object detection [31], and near duplicate detection [32], to mention just a few examples. All of these rely on accurately finding the correspondence(s) of a point on an object in a query image given one or more target images that might contain the same object. In many applications the target images have undergone transformations with respect to the query image; in stereo vision, the viewpoint is different, while in object recognition and near duplicate detection both the lighting and even the object itself may also be transformed.

In the literature two approaches to feature point matching have been pursued and later merged, namely the *geometric approach* and the *descriptor-centric approach*.

In the purely geometric approach, feature points are matched based on their location in the images. [23] and [24] introduced the use of spectral methods by deriving a coherent set of matches from the eigenvalues of the correspondence matrix. Other examples of this approach include [9,22].

The descriptor-centric approach on the other hand finds matches by pairing similar keypoints. The first examples of this approach used the correlation of the raw image data immediately surrounding the feature point [3,12] to calculate this similarity. Later algorithms were enhanced by invariant feature descriptors, as first introduced by [21] and later popularized by the work of [17] introducing SIFT and [4] introducing SURF.

A straightforward way of finding a set of correspondences using only feature points is to apply a threshold to the similarity measure of the feature vectors, accepting only correspondences that score above a certain level of similarity [25]. When we match images with the assumption that the correspondence between two feature points will be unique, we can further increase precision by only matching a feature point to its nearest neighbor in terms of descriptor similarity. Instead of thresholding based on similarity, [12] and [3] proposed using the *ratio* of the similarity of the best to second best correspondence of a given point to evaluate how unique it is. Their finding was later tested by several independent teams, all concluding that thresholding based on this ratio is generally superior to thresholding based on similarity [17–20]. [7] extended this "ratio-match" idea to deal with a set of images by using not the ratio of the best and second best correspondence, but the average ratio of the best and the second best correspondences across a set of images. [20] tried to enhance descriptor matching by looking at the statistical distribution of local features in the matched images, and only return a match when such a correspondence would not occur by mere chance.

☆ This paper has been recommended for acceptance by J.K. Kämäräinen.

* Corresponding author. Tel.: +65 6591 7760; fax: +65 6591 9091.
*E-mail address:* stefan.winkler@adsc.com.sg (S. Winkler).

http://dx.doi.org/10.1016/j.patrec.2015.12.017
0167-8655/

Finally, a precursor of the algorithms discussed in this paper was introduced by the authors as *Mirror-Match*, which makes use of the feature points in both images to decide if a match is valid [2].

A plethora of hybrid solutions have combined descriptor matching with various geometric constraints to improve matching. These constraints are based on assumptions regarding the transformation between query and target images. At the stricter end we have epipolar constraints, assuming that images can be tied by a homography [11,26], and angular constraints, assuming correspondences are angled similarly [14,21]. Often these approaches are made computationally feasible by modeling feature correspondences as an instance of graph matching, where each feature is a vertex, and edge values correspond to a geometric relation between two features. Approximate graph matching algorithms can then be used to efficiently establish an isomorphism between the feature graphs of two images [15,27,29,30]. Others define image regions and reject or accept correspondences based on the regions they connect [10,28].

Any matching method relying on geometric constraints is limited by inherent assumptions about the geometric relationship between the two images. Broad assumptions such as the epipolar constraint only apply in simple image transformations. For more complex transformations we need models suitable for each particular case, which restricts them to the subset of images that fit the model. Transformations from one scene to another often feature a change in perspective, background, and sometimes variations within the object itself: a person can change pose, a car model can have different configurations, a flower can bloom etc. When matching these instances we are forced to either create a sophisticated model that represents the variables of transformation within the object, or alternatively find correspondences using an algorithm with no inherent geometric assumptions. Besides, any geometric method acts as a filter on a given set of correspondences. Therefore, if the initial set of purely descriptor-based matches contains fewer incorrect correspondences, the final set can be calculated faster and more accurately.

The methods we propose in this paper are designed to be free from assumptions about image geometry. They extend and improve on *Ratio-Match* [17] and the authors' *Mirror-Match* [2] by generalizing both algorithms to a framework of matching methods. We go on to formally establish a ranking based on how different methods within the framework compare in terms of precision and recall. Our experimental evaluations confirm the theoretical results and show that *Ratio-Match* is generally a sub-optimal choice as a matching algorithm.

In our previous paper [2], we introduced *Mirror-Match* and *Mirror-Match with Clustering*, two algorithms that outperform the state of the art. The novel contributions of the present paper consist of presenting these algorithms together with several related existing algorithms in a general and comprehensive framework. We further develop the theoretical foundations for comparing the algorithms and use these to formally prove a ranking in terms of performance for the different algorithms. This also enables us to understand why *Mirror-Match* performs better than *Ratio-Match* in the first place. In addition we benchmark all algorithms within the framework extensively on a much larger dataset containing over 3000 image pairs.

The paper is organized as follows. Section 2 introduces the original *Ratio-Match* and extends it to introduce the proposed framework. Section 3 compares the various methods of the framework theoretically. Section 4 presents an experimental evaluation and discusses the results obtained. Section 5 concludes the paper.

## 2. Matching framework

### 2.1. Definitions

The proposed framework is inspired by *Ratio-Match* as introduced by [12] and later used by [3] and [17]. *Ratio-Match* is motivated by the observation that nearest-neighbor feature matching is not necessarily the best strategy [17,18]. The distance between the feature descriptors of two nearest neighbors might tell us on a global level how much they resemble each other, but it does not tell us if other feature points are equally similar. *Ratio-Match* makes use of the ratio between the nearest and second nearest neighbor as a heuristic to determine the confidence of the match. Matches are returned only if this ratio is lower than a given threshold $\tau$, filtering out feature points that are ambiguous because others match almost equally well.

The underlying assumption in *Ratio-Match* is that the point we seek to match in the query image has only one true correspondence in a given target image or no matches at all. In both cases we can infer that the *second* nearest neighbor in the target image is not a true correspondence. We consider the distance between the second nearest neighbor and the feature point as the baseline. It tells us how similar the descriptors of two feature points can be when they are not a true correspondence. Some feature points might have very unique descriptors with large distances to false correspondences, while others may be generic with plenty of similar points. Knowing the baseline for all features allows us to be lenient in the first case and cautious in the second. In practice *Ratio-Match* scores a match by dividing the distance to the nearest neighbor with the distance to the second nearest neighbor (the baseline) to estimate how distinct the correspondence is from a false match.

In what follows we will use the following nomenclature:

- Let $f_q$ be a feature point in the query image.
- Let $\mathcal{F}$ be a set of features. $\mathcal{F}_t$ denotes all features from the target image.
- Let $\tau \in [0 \ldots 1]$ be a threshold used to decide whether to keep a match.
- Let the *proposed match* be the nearest neighbor of a query feature $f_q$ picked from a set of feature points that we call the *proposal set* $\mathcal{F}_p$, which does not contain the query feature.
- Let the *baseline match* be the nearest neighbor of the query feature $f_q$ picked from a set of feature points that we call the *baseline set* $\mathcal{F}_b$, which contains neither the proposed match nor the query feature. In *Ratio-Match* the baseline match is the second nearest neighbor.

### 2.2. Framework of matching methods

We can generalize *Ratio-Match* by expanding on the idea of baseline and proposal sets. With *Ratio-Match* these two sets are created from features in the target image, but this is not the only option. If we use the features in the query image as well as the combined features of both images, we end up with six possible permutations of a *Ratio-Match*-like algorithm. We illustrate these variants in Fig. 2 and will go on to prove theoretically and demonstrate empirically that *Ratio-Match* is among the least performant of the pack.

The algorithms *Ratio-Match*, *Self-Match* and *Both-Match* all find the best match to a given query feature only in the target image. They differ by the feature set used as the baseline set. While *Ratio-Match* uses features from the target image, *Self-Match* draws the baseline set from the query image. Finally *Both-Match* uses the conjunction of features from both images.