Match-time covariance for descriptors

Eric Christiansen[1]
echristiansen@cs.ucsd.edu
Vincent Rabaud[2]
vrabaud@aldebaran-robotics.com
Andrew Ziegler[3]
amz@gatech.edu
David Kriegman[1]
kriegman@cs.ucsd.edu
Serge Belongie[1]
sjb@cs.ucsd.edu

[1] University of California, San Diego
La Jolla, California, USA
[2] Aldebaran Robotics
Paris, France
[3] Georgia Institute of Technology
Atlanta, Georgia, USA

Local descriptor methods are widely used in computer vision to compare local regions of images. These descriptors are often extracted relative to an estimated scale and rotation to provide invariance up to similarity transformations. We call this extract-time covariance (ETC) following the language of [1]. ETC is an imperfect process, however, and can produce errors downstream. Figure 1 illustrates the deterioration of common methods under changing scale and rotation.

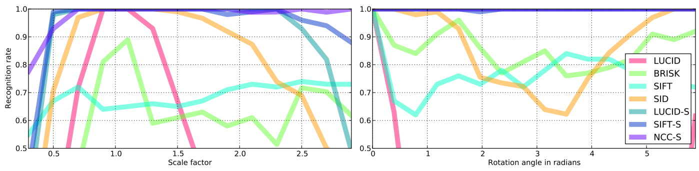

Figure 1: Recognition rates as a function of synthetic scale and rotation of the Oxford `boat` base image for various methods. The rates were obtained following the protocol of [2] and using the 100 strongest keypoints per image. The ETC methods (BRISK and SIFT) pay a penalty for the unreliability of their scale and rotation detectors. The Scale Invariant Descriptor (SID) lacks true scale invariance because it lacks the necessary post-matching normalization. Only the MTC methods (*-S) exhibit true invariance.

In this paper, we propose an alternative to steering we refer to as match-time covariance (MTC). MTC is a general strategy for descriptor design that simultaneously provides invariance in local neighborhood matches together with the associated aligning transformations. We also provide a general framework for endowing existing descriptors with similarity invariance through MTC. The framework, Similarity-MTC, is simple and dramatically improves accuracy. It is illustrated in Figure 2.

Finally, we propose NCC-S, a highly effective descriptor based on classic normalized cross-correlation, designed for fast execution in the Similarity-MTC framework. It is extremely simple, as it is just normalized cross-correlation with a novel normalization scheme, plus some bookkeeping for efficiency. NCC-S is described in Figure 3.

NCC-S is also extremely accurate, dramatically outperforming standard descriptors, which primarily use extract-time covariance. See Figure 4 for comparisons.

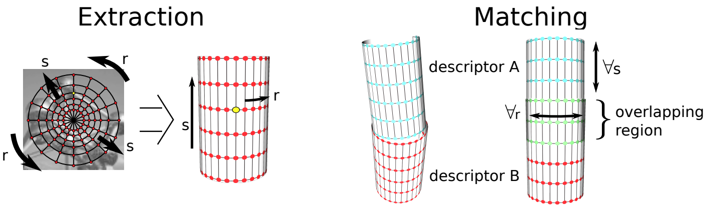

Figure 2: The Similarity-MTC extraction and matching framework, which *wraps* any existing descriptor and provides it with similarity invariance. This description is conceptual; actual implementations may be considerably more efficient. In extraction, descriptors are computed for a range of image scalings and rotations. To do this, a log-polar grid is centered at each keypoint. Each intersection is then associated with the similarity warp which brings the intersection to a fixed canonical location. Taken together, these points describe a cylinder, and descriptor matching is expressed as optimal cylinder alignment. The alignment can be efficiently computed for l_2 descriptor distances.

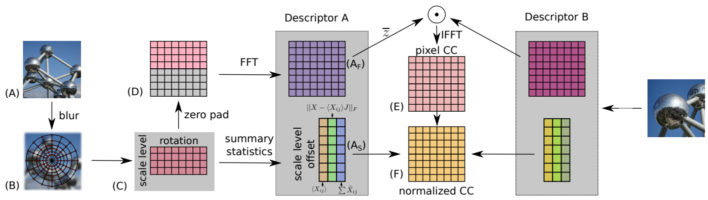

Figure 3: The full pipeline for NCC-S extraction and matching, an efficient descriptor in the Similarity-MTC framework. NCC-S is simply normalized cross-correlation, ported to Similarity-MTC, with some extra tricks for efficiency: 1) It works in Fourier space, 2) It does some bookkeeping to speed normalization.

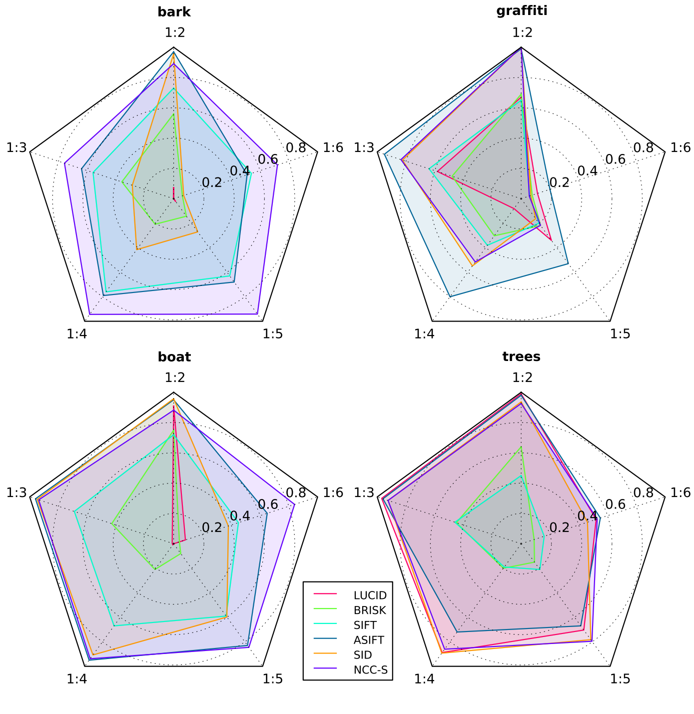

Figure 4: Recognition rates for various methods on a subset of the Oxford image dataset. Each vertex represents an image pair, ranging from `1:2` to `1:6`. The best methods are those that fill their pentagons. NCC-S dramatically outperforms existing methods, with only the much slower ASIFT approaching its accuracy.

The surprising effectiveness of this very simple descriptor suggests that MTC offers fruitful research directions for image matching previously not accessible in the ETC paradigm.

[1] Tinne Tuytelaars and Krystian Mikolajczyk. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 2007.

[2] Andrew Ziegler, Eric Christiansen, David Kriegman, and Serge Belongie. *NIPS*, 2012.