On the Completeness of Coding with Image Features

Wolfgang Förstner
http://www.ipb.uni-bonn.de/foerstner/
Timo Dickscheid
http://www.ipb.uni-bonn.de/timodickscheid/
Falko Schindler
http://www.ipb.uni-bonn.de/falkoschindler/

Department of Photogrammetry
Institute of Geodesy and Geoinformation
University of Bonn
Bonn, Germany

Motivation

Recent local feature detectors extract basic geometric structures like junctions, edges, and dark and bright blobs. The concepts of the detectors are different, sometimes complementary (see Figure 1), in spite of aiming at similar tasks like matching and recognition. Benchmarks usually focus on repeatability and evaluate each method separately [6, 7], without addressing the positive effect when using a combination, which may be useful in many applications [1]. We argue that *completeness* matters in almost any application: The image content should be coded well by the detected features, especially when combining multiple detectors. A tool for measuring completeness is still lacking. We present an evaluation scheme for measuring to what degree local feature detectors cover the image content completely and whether they are complementary in this sense.

Evaluation scheme

We derive a measure d for the incompleteness of a particular set of local features. It expresses the degree to which these features capture relevant image content, which is identified patchwise, over different scales, by the number of bits needed for representing it as in a JPEG coding scheme. Therefore we derive two densities (see Figure 2):

1. An entropy density $p_H(\mathbf{x})$ using the local image statistics. In case the image can be coded with H [bits] we can derive the number of bits per image region $\mathcal{R}$ by $H = \int_{x\in\mathcal{R}} p_H(\mathbf{x})\mathrm{d}\mathbf{x}$.
2. A feature coding density $p_c(\mathbf{x})$. It is derived from a particular set of local features by likewise considering the detection as image coding. Hereby each feature is assumed to be representative for a certain image area, and requiring a certain number of bits for coding.

The two densities are compared. In case p_c is close to p_H the image is efficiently covered with features, and the completeness is high. Thus we require that busy parts of the images are densely covered with features, and smooth parts may not be covered with features.

Entropy density $p_H(\mathbf{x})$. The entropy of a pixel based on coding a patch of size N is denoted by $H(\mathbf{x}, N)$, cf. the paper for its derivation. We determine the expected number of bits per pixel over several scales s by

$$H(\mathbf{x}) = \sum_{s=1}^{S} H(\mathbf{x}, 1+2^s) \tag{1}$$

From this we obtain the entropy density $p_H(\mathbf{x})$ by normalizing $H(\mathbf{x})$ with $\sum_{\mathbf{y}} H(\mathbf{y})$.

Feature coding density $p_c(\mathbf{x})$. Keypoint features f are characterized by their position $\mathbf{m}_f$ and their scale σ_f, or a scale matrix Σ_f, and are usually coded with a fixed number $c(f)$ [bits]. By representing each detected region around a feature with the corresponding Gaussian, we obtain the feature coding map

$$c(\mathbf{x}) = \sum_{f=1}^{F} c(f) G(\mathbf{x}; \mathbf{m}_f, \Sigma_f) \tag{2}$$

We then compute $p_c(\mathbf{x})$ by normalizing $c(\mathbf{x})$ with $\sum_{\mathbf{y}} c(\mathbf{y})$.

Distance measure. We identify the incompleteness d with the distance between $p_H(\mathbf{x})$ and $p_c(\mathbf{x})$, which we express by the Hellinger metric

$$d(p_H(\mathbf{x}), p_c(\mathbf{x})) = \sqrt{\frac{1}{2}\sum_{\mathbf{x}}\left(\sqrt{p_H(\mathbf{x})} - \sqrt{p_c(\mathbf{x})}\right)^2} \tag{3}$$

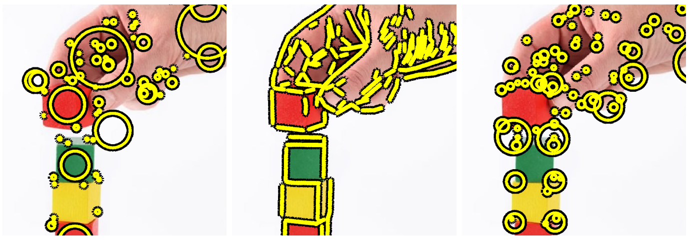

Figure 1: Complementary feature sets, capturing different information of an image. From left: SFOP junctions [2], edges [3], Laplacian blobs [4].

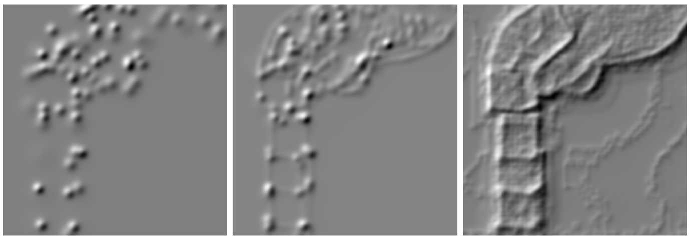

Figure 2: From left: Feature coding densities p_c for SFOP junctions and for the set of three detectors, and estimated entropy distribution p_H.

Results

We evaluated combinations of representative region and keypoint detectors, including Lowe [4], MSER [5] and scale invariant junction features from the recently published SFOP [2] detector, on popular datasets. Our main observations are:

1. Affine covariant detectors do not seem to systematically improve completeness compared to scale and rotation invariant detectors.
2. The classical LOWE detector is most efficiently complemented by the SFOP detector and – almost comparable – the MSER detector.
3. A very good and stable complement is achieved when combining Lowe, SFOP junctions and MSER.

The procedure may well be used in addition to application specific evaluations to characterize the output of future detectors *w.r.t.* existing ones.

[1] T. Dickscheid and W. Förstner. Evaluating the Suitability of Feature Detectors for Automatic Image Orientation Systems. In *International Conference on Computer Vision Systems*, Liege, Belgium, 2009.

[2] W. Förstner, T. Dickscheid, and F. Schindler. Detecting Interpretable and Accurate Scale-Invariant Keypoints. In *ICCV*, Kyoto, Japan, 2009.

[3] Wolfgang Förstner. A Framework for Low Level Feature Extraction. In *ECCV*, pages 383–394, Stockholm, Sweden, 1994.

[4] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004.

[5] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. *Image and Vision Computing*, 22:761–767, 2004.

[6] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and Luc Van Gool. A Comparison of Affine Region Detectors. *IJCV*, 65(1/2):43–72, 2005.

[7] P. Moreels and P. Perona. Evaluation of Features Detectors and Descriptors based on 3D objects. In *IJCV*, 2006.