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
Inst. of Geodesy and Geoinformation
University of Bonn
Bonn, Germany

Abstract

We present a scheme for measuring completeness of local feature extraction in terms of image coding. Completeness is here considered as good coverage of relevant image information by the features. As each feature requires a certain number of bits which are representative for a certain subregion of the image, we interpret the coverage as a sparse coding scheme. The measure is therefore based on a comparison of two densities over the image domain: An entropy density $p_H(\mathbf{x})$ based on local image statistics, and a feature coding density $p_c(\mathbf{x})$ which is directly computed from each particular set of local features. Motivated by the coding scheme in JPEG, the entropy distribution is derived from the power spectrum of local patches around each pixel position in a statistically sound manner. As the total number of bits for coding the image and for representing it with local features may be different, we measure incompleteness by the Hellinger distance between $p_H(\mathbf{x})$ and $p_c(\mathbf{x})$. We will derive a procedure for measuring incompleteness of possibly mixed sets of local features and show results on standard datasets using some of the most popular region and keypoint detectors, including Lowe, MSER and the recently published SFOP detectors. Furthermore, we will draw some interesting conclusions about the complementarity of detectors.

1 Introduction

Local image features play a crucial role in many computer vision tasks, most importantly as an input for camera calibration and object recognition. The basic idea of using local features is to represent the image content by small, possibly overlapping, independent parts which are robust to a number of image distortions up to varying degrees. By identifying such parts in different images of the same scene or object, it becomes computationally feasible to make reliable statements about both image geometry and scene content.

Many algorithms for detecting salient and stable features are available, providing a range of alternative methods with different definitions of *salience* and *stability*. Among the different detectors are methods for extracting basic geometric structures like junctions, circles and edges, and methods for extracting dark and bright blobs. Whilst the particular requirements

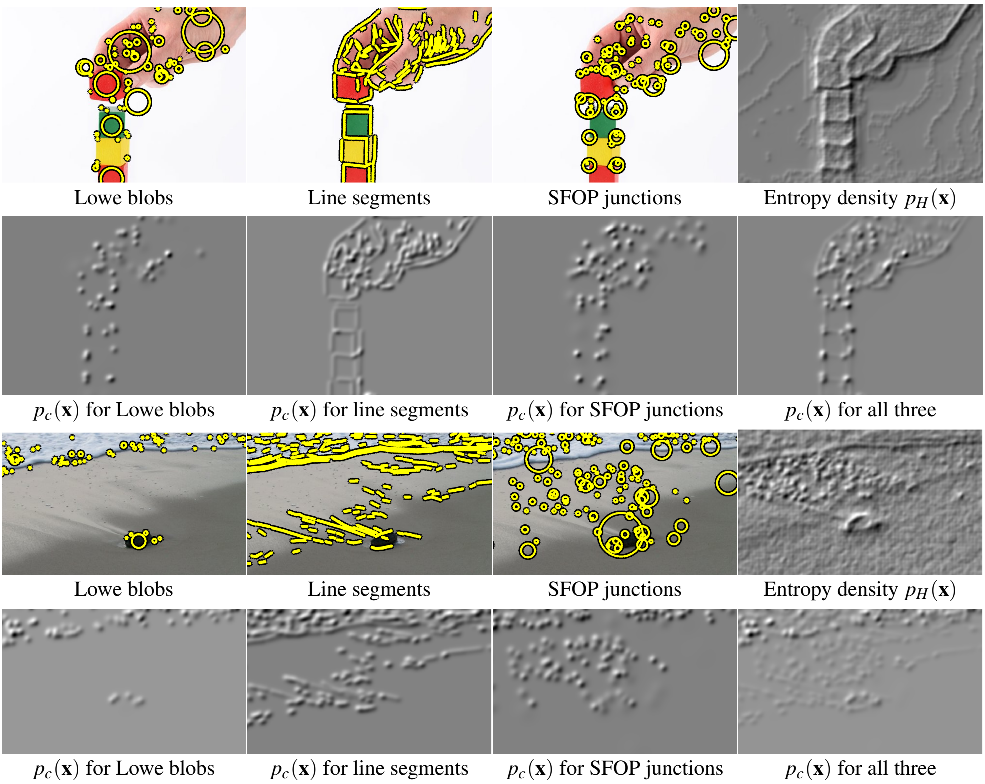

Figure 1: Sets of extracted features on example images *Bricks* and *Beach*, capturing different types and amount of image information, and corresponding entropy densities $p_H(\mathbf{x})$ and feature coding densities $p_c(\mathbf{x})$. We illustrate Lowe blobs [17], straight edge segments [6], SFOP junction features [5], using $\alpha = 0$ in eq. (7)].

for saliency and stability vary strongly depending on the application, we argue that *completeness* of a feature detector is mostly application independent: The image content should be coded well by the detected features.

Let us illustrate the idea on the two simple images depicted in Figure 1. In the top row, the detector by Lowe [17] seems to cover much of the areas representing the visible objects. However, the characteristic contours and properties of the bricks are better coded by the SFOP junctions [5] and especially the edge segments. For coding all relevant parts of the image, one would probably decide to use all three detectors. Indeed, the three exemplary detectors chosen here are highly complementary by design. In the third row of Figure 1, it obviously becomes more difficult to code the whole image content with local features. We see that the Lowe detector covers the piece of rock quite well and also captures the roughness of the surge, but fails in coding the borders between water and dry and wet sand, respectively. The SFOP junctions are more focused on the discriminant parts of such contours, but also clearly lack the completeness of contour coverage provided by the edges.

The major part of this contribution consists in deriving a useful measure d for the incompleteness of a particular set of local features in terms of image coding. We require d

to take small values if a feature set covers relevant image content in a similar manner as a good compression algorithm would. Conversely d should take increasing values for features covering unimportant content as well as for relevant content which is not covered. Therefore we propose to compute two densities over the image domain, both of which are also illustrated in Figure 1: An entropy density $p_H(\mathbf{x})$ based on local image statistics, and a feature coding density $p_c(\mathbf{x})$ which is directly computed from each particular set of local features. The incompleteness is determined as the distance d between $p_H(\mathbf{x})$ and $p_c(\mathbf{x})$, namely the Hellinger metric.

After a short overview on the related work in section 2 we will give a derivation of both densities in section 3 together with an explanation of the proposed scheme. In order to show the feasibility and practical relevance of the scheme, we use various feature detectors and well-known image datasets, as explained in section 4. We finally conclude with a brief summary in section 6.

2 Related Work

Local feature detectors A broad range of local feature detectors with quite different properties is available today. We will shortly introduce some of them here and refer to the detailed review given in [27]. The scale invariant blob detector proposed by Lowe [17] is by far the most prominent one. As it is based on finding local extrema of the response of the Laplacian, which has the well known Mexican hat form, it conceptually aims at extracting dark and bright blobs on characteristic scales of an image. The Hessian affine detector introduced by Mikolajczyk and Schmid [20] based on the theory of Lindeberg [15] also relies on the second derivatives of the image function over scale space. Some detectors directly determine image regions that are characterized by their boundary, such as the intensity- and edge-based region detectors by Tuytelaars and Van Gool [28] and the Maximally Stable Extremal Regions (MSER) by Matas et al. [19].

Another family of detectors is based on the second moment matrix, or structure tensor, computed from the squared gradients of the image functions. It has initially been used by the classical detectors of Harris and Stephens [10] and Förstner and Gülch [7]. Two schemes for exploiting the structure tensor over scale space have been proposed: In [20], scale is still determined by investigating the second derivatives, but the location is determined using the structure tensor. Lindeberg [15] uses the junction model of [7], determines the differentiation scale by analyzing the local sharpness of the image and chooses the integration scale by optimizing the precision of junction localization. The structure tensor may also be utilized for extracting edge segments, as for example in [6]. We do not detail different methods for detecting line segments here and refer to Heath et al. [11], for example.

The image content covered by such detectors, and hence the proportion of information that is thereby coded, varies strongly depending on the type of detector, as depicted in Figure 1. This variation is natural, as pointed out by Triggs [26] who stated that there is no such thing as a generic keypoint detector. The concepts of the detectors are different, sometimes complementary, in spite of aiming at the same tasks, mainly matching and recognition.

There has been high effort in comparing the performance of local feature detectors. A generally accepted criterion is the repeatability of extracted patches under viewpoint and illumination changes in mostly planar scenes [21] and on 3D structured objects [22], but also localization accuracy [9, 29] and the general impact on automatic bundle adjustment [3] has been studied.

Benchmarks usually evaluate each method separately, not addressing the positive effect when using a combination of multiple detectors, which may be very useful in many applications [3]. For example, Bay et al. [1] propose to use edge segments together with regions for calibrating images of man-made environments with poor texture. We experienced a surprisingly strong benefit when using mostly "complementary" combinations of detectors in the framework of [1]. A tool for measuring this complementarity however is still lacking.

The intent of this contribution is to develop an evaluation scheme for measuring in how far the detectors cover the image content completely and whether they are complementary in this sense.

Image statistics For evaluating the completeness of detectors, the question arises which kind of complete but yet sparse coding of an image we take as a reference. The classical, biologically inspired view of Marr [18] is to extract the so-called *primal sketch*, which mainly refers to the blobs, edges, ends and zero-crossings detectable in an image, and is hence achieved to a certain extend by using a combined set of different feature detectors. Marr's approach has been supported by the more recent work of Olshausen and Field [24]. An excellent but yet compact insight into such aspects of image statistics can be found in Mumford [23].

Another approach is motivated from information theory: If we split the image into patches of equal size, a complete but preferably sparse coding would focus on patches with high information content, measured in terms of entropy in the sense of Shannon [25], where entropy is defined as the average (or expected) amount of information transmitted when transferring a signal from sender to receiver.

The entropy of a signal depends on the probability distribution of the events, which is usually not known and hence has to be estimated from data. For example, for pixels in local image patches, the probability distribution may well be modeled by Gaussian densities representing each pixel intensity in the patch. The correlations between pixels, however, are not known a priori and difficult to estimate, so the joint probability distribution is unknown. J.-F. Bercher [12] proposed a general method for estimating the entropy of a signal by modeling the unknown distribution as an autoregression process, focusing on finding a good approximation with tractable numerical integration properties.

The widely accepted JPEG image coding scheme is also based on entropy encoding. To account for the unknown pixel correlations, the data compression is carried out after a Discrete Cosine Transformation (DCT) of the image, computed on local 8×8 patches or on a wavelet transform of the complete image. The coefficients are then coded in blocks. A comparison of image coding schemes is given in Davis and Nosratinia [2], for example. Grazzini et al. [8] found that maxima of local entropy play a major part for extracting singular manifolds, that is statistically important parts, in infrared satellite images. They also addressed the problem of scale, and proposed to use larger patch sizes and weight pixels according to their distance to the center pixel before computing the spectrum.

In contrast to using image statistics as a prior for image analysis, we aim at taking a certain aspect only, namely the local entropy as a reference for evaluating the ability of feature detectors to completely cover the image.

3 Theory

Local feature detection can be considered as coding of image content. Measuring completeness of local feature extraction may consequently be considered as a comparison with classical coding schemes. This requires to develop

1. an entropy density $p_H(\mathbf{x})$ using local image statistics. If 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 detected local features, assuming each feature is representative for a certain image area.

The two densities can then be compared. In case p_c is close to p_H, the image is efficiently covered with features, and the completeness is high. We hereby require busy parts of images to be densely covered with features, and smooth parts not to be covered with features.

Furthermore, we have to take into account that many detectors are able to find features of different scales at the same image position. This is easy to realize regarding p_c, as each feature represents a certain image area, thus the bits for coding the image feature are spread over that area. For the entropy density p_H, we assume that the coding is hierarchical, like in wavelet transform coding where information from different scales is integrated.

3.1 Entropy density

We determine the entropy density p_H based on small image patches of different size, or scale, respectively. Given a square image patch $g(i,j), i,j = 1,...,N$, we only have one sample of the distribution of image content in that patch. We therefore assume the image patch to be representative for a large image. Specifically, we assume that it is a subsection of an infinite doubly periodic image with period N in both directions, allowing us to represent it in the frequency domain. Furthermore we assume the image to be the noisy version of a Gaussian process.

We can derive the entropy from the covariance matrix Σ_{gg} of the N^2 intensity values, which itself can be derived from the autocorrelation function and noise variance σ_n^2:

$$H(\mathbf{g}) = \frac{1}{2}\log_2 2\pi \mathrm{e}\frac{|\Sigma_{gg}|}{\sigma_n^2} \quad \text{[bits/patch]} \tag{1}$$

Due to Parseval's identity, the determinant $|\Sigma_{gg}|$ can easily be derived from the power spectrum $P(\mathbf{u}) = |\mathrm{DCT}(g(\mathbf{x}))|^2$ via $|\Sigma_{gg}| = \prod_{\mathbf{u}\backslash\mathbf{0}} P(u,v)$. Note that we omit the DC term $P(\mathbf{0})$, as the covariance matrix captures only deviations from the mean. We use the Discrete Cosine Transform (DCT) instead of the Fast Fourier Transform in order to reduce the effects of patch borders. We assume the powerspectrum to be additively composed of the powerspectra of the signal and the noise. Thus it theoretically is limited from below by the noise variance σ_n^2 of the graylevels. We therefore use the regularized estimate for the power spectrum of the signal

$$\widehat{P}(\mathbf{u}) = \max\left(P(\mathbf{u}) - \sigma_n^2, 0\right) . \tag{2}$$

We then determine the entropy of a single pixel in the image patch from

$$H(g) = \frac{1}{2N^2}\sum_{\mathbf{u}\backslash\mathbf{0}} \max\left(\log_2 2\pi\mathrm{e}\frac{\max\left(P(\mathbf{u})-\sigma_n^2, 0\right)}{\sigma_n^2}, 0\right) \quad \text{[bits/pixel]} . \tag{3}$$

excluding frequencies where the entropy would be negative. We assume the noise to be larger than the rounding error introduced by graylevel quantization, thus $\sigma_n \geq \varepsilon/\sqrt{12} \approx 0.29$ [gr], where ε is the unit of the graylevels. This setup guarantees the entropy to be $H(g) \geq 0$.

We now integrate the expected number of bits per pixel over scales. Let the entropy of a pixel $\mathbf{x}$ based on patch size N be $H(\mathbf{x}, N)$. As a first choice, we determine the expected number of bits per pixel over several scales by the equally weighted sum

$$H(\mathbf{x}) = \sum_{s=1}^{S} H(\mathbf{x}, 1+2^s) \tag{4}$$

We still have to investigate whether this is an optimal choice. In our experiments we use $S = 7$, thus the patch size is limited to $3 \leq N \leq 129$. Observe that omitting the DC term in a lower scale is compensated by the coding in the higher scales.

Finally we obtain the entropy density by normalizing $H(\mathbf{x})$:

$$p_H(\mathbf{x}) = \frac{H(\mathbf{x})}{\sum_{\mathbf{y}} H(\mathbf{y})} \tag{5}$$

The expected number of bits in a certain region $\mathcal{R}$ therefore is $H \sum_{\mathbf{x}\in\mathcal{R}} p_H(\mathbf{x})$, where H is the total number of bits for the complete image. However, for comparing the entropy distribution over the image with a particular set of local features, we can not use absolute values (*i.e.* bits per pixel), but only relative values. This is why we take $p_H(\mathbf{x})$ as reference. In the right column of Figure 1, the density $p_H(\mathbf{x})$ is depicted for two example images.

3.2 Feature coding density

Feature detectors usually deliver sets of features which are representative for a certain region. For each feature type we need a certain amount of bits coding the respective region. Let a feature f require $c(f)$ [bits], being responsible for a region $\mathcal{R}(f)$. Assuming a uniform density of the bits per pixel within each region, we obtain a feature coding map from all F features

$$c(\mathbf{x}) = \sum_{f=1}^{F} \frac{\mathbf{1}_{\mathcal{R}(f)}(\mathbf{x})}{|\mathcal{R}(f)|} c(f) \tag{6}$$

where $\mathbf{1}_{\mathcal{R}(f)}(\mathbf{x})$ is an indicator function, being 1 within the region $\mathcal{R}(f)$, and 0 outside.

In our case we analyse keypoint features f which are characterized by their position $\mathbf{m}_f$ and their scale σ_f, or a scale matrix Σ_f in case of affine invariant features. In case we analyse straight edge segments $(\mathbf{x}_S, \mathbf{x}_E)_f$ we use $\mathbf{m}_f = (\mathbf{x}_{S,f} + \mathbf{x}_{E,f})/2$ and a covariance matrix with major semi-axis $|\mathbf{x}_{E,f} - \mathbf{x}_{S,f}|/2$ in the direction of the edge and minor semi-axis $\sigma_a = 1$ [pel]. Therefore it is reasonable to replace the uniform distribution by a Gaussian and derive the feature coding map by

$$c(\mathbf{x}) = \sum_{f=1}^{F} c(f) G(\mathbf{x}; \mathbf{m}_f, \Sigma_f) \tag{7}$$

The actual density to be compared with the entropy density $p_H(\mathbf{x})$ is then

$$p_c(\mathbf{x}) = \frac{c(\mathbf{x})}{\sum_{\mathbf{y}} c(\mathbf{y})}. \tag{8}$$

Figure 2: Example images from the datasets used for the experiments.

3.3 Evaluating completeness of feature detection

In case the empirical feature coding density $p_c(\mathbf{x})$ would be identical to the entropy density $p_H(\mathbf{x})$, coding of the image with features would cover the image in the same manner as using image compression. Of course image features may use less or more bits for coding the complete image, depending on the coding of the individual feature, so we do not compare the absolute number of bits per pixel, but their densities. We use Hellinger's metric $d(p(x), q(x))$ of two densities $p(x)$ and $q(x)$ for measuring the difference between $p_H(\mathbf{x})$ and $p_c(\mathbf{x})$:

$$d(p_H(\mathbf{x}), p_c(\mathbf{x})) = \sqrt{\frac{1}{2}\sum_{\mathbf{x}}\left(\sqrt{p_H(\mathbf{x})} - \sqrt{p_c(\mathbf{x})}\right)^2} \tag{9}$$

4 Experiments

Applied feature detectors. For providing a good spectrum of the available methods, we selected most of the prominent feature detectors presented in Mikolajczyk et al. [21], including both the non-affine and affine versions of the scale-invariant Harris and Hessian detectors from [20]. The implementations are taken from the website maintained by the authors of [21]. For the Lowe detector, we used the original sourcecode kindly provided by the author, but starting with the original instead of the double image resolution for comparability. Additionally, we used scale invariant SFOP junction features [5], and a classical edge detector based on the theory in [6], which uses a minimum edge length of ten pixels. The SFOP junctions are obtained as a subset by restricting to $\alpha = 0$ [5, eq. (7)].

Image data. Our objective is to compare the completeness of these feature detectors on a wide variety of images. For that purpose, we used the fifteen natural scene category dataset [13, 14] and added the well-known Brodatz texture dataset, a collection of cartoon images and a set of subimages of an aerial image as further categories. We present here results for a subset of these datasets only, as depicted by some example images in Figure 2. Our main conclusions however hold for all datasets.

Experimental setup. We compute the incompleteness measure d for each of the images and each of the detectors mentioned above. As we are interested in the effect of combining different features, especially for finding evidence about their complementarity, we also consider combinations of features. However, as evaluating all possible combinations is intractable within the scope of this paper, we focus on the potential of feature detectors to complement the Lowe detector for now.

Clearly, completeness raises when the significance level of a detector is lowered, which conversely reduces repeatability and performance in practice. We used the default settings

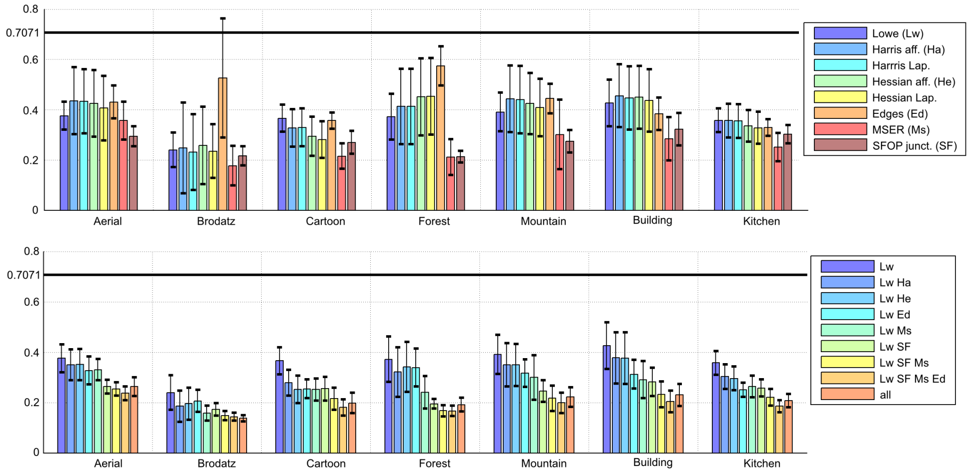

Figure 3: Experimental results for separate (top row) and combined (bottom row) feature detectors. Abbreviations used in the bottom legend are introduced in the top legend.

as applied in the existing benchmarks for the standard detectors, and kept the settings for the SFOP detector so that the number of features is comparable to that of Lowe. It is important to note that too low significance levels would distort the results, as noisy features on homogeneous areas would increase the proposed metric. In order to minimize such effects, the actual image noise should be taken into account, either by noise removal in the image or consideration of the estimated noise in the algorithm [4, 16].

5 Results

Figure 3 shows the average difference d between the entropy density and the feature coding density for each operator and seven image categories. In the plots, the mean value over all images of an image category is given, together with the 1-σ-confidence region, denoted by the vertical black bars. Note that this is the standard deviation of the samples, not the the standard deviation of the mean values, which would clearly be smaller. The distance d is related to the angle between the vectors $\sqrt{p_H}$ and $\sqrt{p_c}$. We have plotted a boundary at $d_0 = \sqrt{1}/\sqrt{2}$ which corresponds to an angle of $90°$, indicating a threshold beyond which we treat densities as sufficiently different. The value d_0 is indicated by a level line in both diagrams.

5.1 Results for separate detectors

The average completeness per image category is quite similar for the Lowe, Harris affine and Hessian affine detectors: It ranges around 0.37 to 0.39, with a significantly better result on the *Brodatz* dataset (~ 0.25) and worst results for *Mountain* and *Tall Buildings*. The standard deviation however is significantly lower for the Lowe detector than for the other two. Higher completeness *w.r.t.* the *Brodatz* images can be observed for all other detectors as well, which can be understood from the fact that these pictures possess a strong uniform structuredness.

The SFOP junctions show significantly higher completeness in all categories than the previous three. The distance measure on average is more than 25% smaller. The standard deviation is comparable to that of Lowe.

The MSER detector achieves best overall completeness as a single detector for *Brodatz* and *Forest*, although on average the results are comparable with that of the SFOP junctions. Compared to its overall results, MSER has a quite bad completeness for *Aerial*, and - due to the similarity of the texture - also for *Mountain*. Edge features alone show a rather bad completeness in most cases, which may be explained from the fact that on most images, only the boundaries of informative image regions are extracted. On datasets dominated by visible contours like *Kitchen* and *Building* however, coverage by edges is at least average.

In the upper plot of Figure 3 we also show results for the Harris and Hessian affine detectors compared to the so called Harris- and Hessian-Laplace detectors [20]. For the latter ones, the final estimation of affine parameters for the local pixel neighbourhood is left out, so that the resulting patches are circular instead of elliptical. We expected better results for the affine versions, which has not been confirmed by the experiments.

5.2 Results for combinations of features

Our expectation on the results for pairwise combinations was the following: Combining the Lowe with the Hessian affine detector should not yield significant improvement, as both use the Laplacian to identify the optimal scale. For MSER we expect stronger improvements, as it looks for homogeneous regions. The combination with the Harris and SFOP junction detector should give the best completeness, as they rely on the structure tensor and hence search for points with many gradients instead of bright or dark blobs.

The results for pairwise combinations are shown left in the bottom row of Figure 3. The improvement achieved by combining Lowe with Hessian affine features is only slight, *i.e.* below 10% except for *Brodatz* and *Cartoon*. Other than expected, the Harris affine detector complements Lowe no better than Hessian affine, which is very surprising. A possible reason for this observation is that Harris affine is no “pure” junction detector: Characteristic scales for detected junction points are searched in the Laplacian scale space. Moreover, the Harris affine detector is known to find multiple keypoints with overlapping regions at junctions.

The SFOP junction detector seems to be the best pairwise complement to the Lowe detector. Here the improvement on the Lowe detector alone is at least 10%, often over 20%. This becomes most obvious for *Mountain*. The MSER detector complements Lowe better than Harris and Hessian, but not as significant as the SFOP junctions.

Using more different detectors at once, we get slight improvements with MSER upon the pairwise combination of Lowe and SFOP junctions, as can be seen from the yellow bars in the bottom row of Figure 3. The result of these three seems to be optimal in our experiments: it is not noticeably changed when adding more detectors.

6 Conclusion and outlook

We have proposed an intuitive scheme for measuring completeness of local features in the sense of image coding. To achieve this, we derived suitable estimates for the distribution of relevant image information and the coverage by a set of local features over the image domain, which we compared by the Hellinger's distance. This enables us to analyze the completeness of different sets of local features over arbitrary image sets.

We made a number of interesting observations: Affine detectors do not seem to systematically improve completeness compared to scale and rotation invariant detectors.

The Harris affine detector shows significantly lower complementarity *w.r.t.* the Lowe detector than the SFOP junctions. It actually behaved very similar to the Hessian affine detector in our experiments. Probably the fact that the Harris affine detector uses the Laplacian for scale selection is one reason for this. The SFOP junctions complement the Lowe detector most significantly, while the effect for MSER is only slightly worse. A very good and stable complement is achieved when combining Lowe blobs, MSER regions and SFOP junctions.

The proposed entropy density $p_H(\mathbf{x})$ gives rise to a new scale invariant keypoint detector which locally maximizes the entropy over position and scale. We will analyse its completeness *w.r.t.* image coding, similar as proposed in this paper, and its applicability as a keypoint detector for matching and object detection.

References

[1] Herbert Bay, Vittorio Ferrari, and Luc Van Gool. Wide-Baseline Stereo Matching with Line Segments. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 329–336, Washington, DC, USA, 2005.

[2] Geoffrey M. Davis and Aria Nosratinia. Wavelet-Based Image Coding: An Overview. *Applied and Computational Control, Signals, and Circuits*, 1:205–269, 1998.

[3] T. Dickscheid and W. Förstner. Evaluating the Suitability of Feature Detectors for Automatic Image Orientation Systems. In *7th International Conference on Computer Vision Systems (ICVS'09)*, Liege, Belgium, 2009.

[4] W. Förstner. Image Preprocessing for Feature Extraction in Digital Intensity, Color and Range Images. In *Geomatic Methods for the Analysis of Data in Earth Sciences*, volume 95/2000 of *Lecture Notes in Earth Sciences*, pages 165–189. Springer, 1998.

[5] W. Förstner, T. Dickscheid, and F. Schindler. Detecting Interpretable and Accurate Scale-Invariant Keypoints. In *12th IEEE International Conference on Computer Vision (ICCV'09)*, Kyoto, Japan, 2009.

[6] Wolfgang Förstner. A Framework for Low Level Feature Extraction. In *Third European Conference on Computer Vision (ECCV'94)*, volume III, pages 383–394, Stockholm, Sweden, 1994. ISBN 0-387-57957-5.

[7] Wolfgang Förstner and Eberhard Gülch. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. In *ISPRS Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, June 1987.

[8] Jacopo Grazzini, Antonio Turiel, and Hussein Yahia. Entropy Estimation and Multiscale Processing in Meteorological Satellite Images. *International Conference on Pattern Recognition (ICPR'02)*, 3:30764, 2002.

[9] A. Haja, B. Jähne, and S. Abraham. Localization Accuracy of Region Detectors. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08)*, June 2008. ISSN 1063-6919.

[10] C. Harris and M. J. Stephens. A Combined Corner and Edge Detector. In *Alvey Vision Conference*, pages 147–152, 1988.

[11] Mike Heath, Sudeep Sarkar, Thomas Sanocki, and Kevin Bowyer. Comparison of Edge Detectors: A Methodology and Initial Study. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'96)*, 0:143, 1996.

[12] C. Vignat J.-F. Bercher. Estimating the Entropy of a Signal with Applications. *IEEE Transactions on Signal Processing*, 48(6), June 2000.

[13] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pages 2169–2178, Washington, DC, USA, 2006.

[14] Fei-Fei Li and Pietro Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 524–531, Washington, DC, USA, 2005. IEEE Computer Society.

[15] Tony Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[16] Ce Liu, Richard Szeliski, Sing Bing Kang, C. Lawrence Zitnick, and William T. Freeman. Automatic estimation and removal of noise from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):299–314, 2008.

[17] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[18] David Marr. *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco, 1982.

[19] 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.

[20] Krystian Mikolajczyk and Cordelia Schmid. Scale and Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

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

[22] P. Moreels and P. Perona. Evaluation of Features Detectors and Descriptors based on 3D objects. In *International Journal of Computer Vision*, 2006.

[23] David Mumford. Empirical Statistics and Stochastic Models for Visual Signals. In S. Haykin, J.C. Principe, T.J. Sejnowski, and J. McWhirter, editors, *New Directions in Statistical Signal Processing: From Systems to Brain*. MIT Press, Cambridge, MA, 2005.

[24] B. A. Olshausen and D. J. Field. Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature*, 381(6583):607–609, June 1996.

[25] Claude E. Shannon. A Mathematical Theory of Communication. Technical report, Bell System Technical Journal, 1948.

[26] Bill Triggs. Detecting Keypoints with Stable Position, Orientation, and Scale under Illumination Changes. In *8th European Conference on Computer Vision (ECCV'04)*, pages 100–113, 2004.

[27] Tinne Tuytelaars and Krystian Mikolajczyk. *Local Invariant Feature Detectors: A Survey*. Now Publishers Inc., Hanover, MA, USA, 2008.

[28] Tinne Tuytelaars and Luc Van Gool. Matching Widely Separated Views Based on Affine Invariant Regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.

[29] B. Zeisl, P. Georgel, F. Schweiger, E.Steinbach, and N. Navab. Estimation of Location Uncertainty for Scale Invariant Feature Points. In *British Machine Vision Conference*, London, UK, 2009.