Estimation of Location Uncertainty for Scale Invariant Feature Points

Bernhard Zeisl[1]
http://campar.in.tum.de/Main/BernhardZeisl
Pierre Fite Georgel[1]
http://campar.in.tum.de/Main/PierreGeorgel
Florian Schweiger[2]
http://www.lmt.ei.tum.de/team/florian/
Eckehard Steinbach[2]
http://www.lmt.ei.tum.de/team/steinb/
Nassir Navab[1]
http://campar.in.tum.de/Main/NassirNavab

[1] Chair for Computer Aided Medical Procedures & Augmented Reality
Technische Universität München
Munich, GER

[2] Institute for Media Technology
Technische Universität München,
Munich, GER

State of the art algorithms detect features that are invariant to scale and orientation changes. While feature detectors and descriptors have been widely studied in terms of stability and repeatability [4, 6], their localisation error has often been assumed to be uniform and insignificant. We argue that this assumption does not hold for scale-invariant feature detectors and demonstrate that the detection of features at different image scales actually has an influence on the localisation accuracy.

Uncertainty estimation for corner-like points, which are *not* scale-invariant, as a measure for the localisation precision has been studied before. Common to all approaches is the assumption of a Gaussian error model and hence the characterization using a 2D anisotropic covariance matrix with varying orientation and magnitude. Considering the curvature of the self-matching residual at a feature point, this covariance can be estimated from the second moment matrix [1, 3]. Another approach is to propagate covariances of a noise model for pixel intensities through the detection process [7]. [5] evaluates the accuracy for Harris corner points, while [2] looks at the matching precision of interest regions.

Compared to corner detectors, scale-invariant region detectors extract image regions complementary to corner-like features. Hence we claim two things: *First, due to the focus on interest regions, the shape of covariances will be in general anisotropic. Second, the magnitude of covariances will vary significantly due to detection in scale space.* We introduce a general framework to determine the uncertainty of multi-scale image features. The framework is applied to the well-known SIFT and SURF algorithms, and we detail its implementation and make it available [1].

Our analysis is based on the assumption that the detection process localises a feature and generates a measurement error that conforms to a bivariate normal distribution. Common to all scale invariant feature detectors is a two step approach to find feature points. First, a scale-space representation of the detector response in form of an image stack D is created. Within this stack, extrema relate to feature points $\langle \mathbf{p}, \sigma_i \rangle$:

$$D(\mathbf{x}, \sigma_i) = \mathbf{f_{dec}}\left(I(\mathcal{N}_\mathbf{x}), \sigma_i\right) \tag{1}$$

$$\mathbb{P}_1 := \bigcup_{i=1}^{N} \left\{ \langle \mathbf{p}, \sigma_i \rangle \,\middle|\, \mathbf{p} = \arg\max_{\mathbf{x} \in \mathcal{N}_\mathbf{p}} \left(D(\mathbf{x}, \sigma_i)\right) \right\} \tag{2}$$

Second, the algorithm selects for each feature point its characteristic scale σ, i.e., the scale at which a maximum detector response to the local image structure is observed.

For the evaluation of a localisation error only the feature detection operator and by this means the particular layer $D(\bullet, \sigma)$ of the detection stack is the determining factor. Given a small neighborhood $\Delta\mathbf{p}$ around a feature point location $\mathbf{p}$ we can approximate the detector response map D via a Taylor expansion up to second order with the following residual.

$$R(\Delta\mathbf{p}) = |D(\mathbf{p}, \sigma) - D(\mathbf{p} + \Delta\mathbf{p}, \sigma)| \approx \frac{1}{2} \Delta\mathbf{p}^\top \mathbf{H} \Delta\mathbf{p} \tag{3}$$

The Hessian $\mathbf{H}$ describes the curvature at the feature point $\mathbf{p}$. Simply speaking *for a low curvature the detection process will be error prone due to the missing discriminative behaviour of D in the neighborhood of* $\mathbf{p}$, *whereas for a high curvature the spacial detection process will be more accurate*. Following the argumentation in [3] we take the inverse of $\mathbf{H}$ as our covariance estimate for each individual feature point $\langle \mathbf{p}, \sigma \rangle$:

$$\Sigma = \mathbf{H}^{-1} = \begin{bmatrix} R_{xx}(\mathbf{p}, \sigma) & R_{xy}(\mathbf{p}, \sigma) \\ R_{xy}(\mathbf{p}, \sigma) & R_{yy}(\mathbf{p}, \sigma) \end{bmatrix}^{-1} \tag{4}$$

[1] Binaries/code at `http://campar.in.tum.de/Main/CovarianceEstimator`

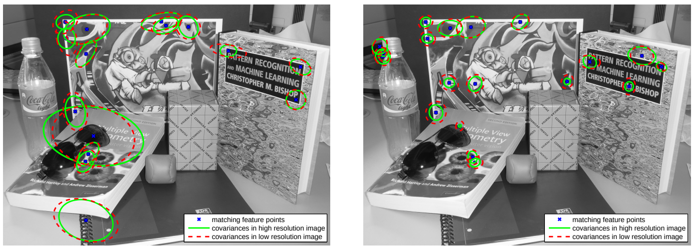

Figure 1: Covariances for matching feature points between a low and high resolution image; left SIFT, right SURF. Covariances from the low resolution image are projected to the high resolution image.

Within a synthetic setup (where we are able to control the ground truth feature point location) experiments with the SIFT and SURF feature detectors show that our covariance estimates represent the underlying localisation error distribution. As assumed, features detected at higher scales are less accurate compared to features detected at lower scales.

The effect of greater localisation error for larger image features can e.g. be utilized in registering two differently sized images. If the two images are of different size, the error minimization between matching feature points will mostly depend on the error in the higher resolution image. Incorporating our covariances achieves an *automatic* error normalization, such that both images account equally to the cost function (see Figure 1).

We also apply our covariances in model fitting algorithms, which minimize a least squares problem. Considering covariances, minimisation of the Euclidian distance results in minimising the Mahalanobis distance. Thereby, terms with large covariances maintain less influence on the overall cost reduction, resulting in a weighted least square optimization. For bundle adjustment we were able to show a decrease in reprojection error in our setup from 2.03 to 1.76 pixel for SIFT and 2.55 to 2.36 pixel for SURF.

Acknowledgment We would like to thank Marc Pollefeys for the inspiring discussion about the normalising behaviour of covariances for differently sized images.

[1] M.J. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel. What value covariance information in estimating vision parameters? In *Proc. ICCV*, volume 1, 2001.

[2] A. Haja, B. Jahne, and S. Abraham. Localization accuracy of region detectors. In *Proc. CVPR*, pages 1–8, 2008.

[3] Y. Kanazawa and K. Kanatani. Do we really have to consider covariance matrices for image features? In *Proc. ICCV*, volume 2, 2001.

[4] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool. A comparison of affine region detectors. *IJCV*, 65(1):43–72, 2005.

[5] U. Orguner and F. Gustafsson. Statistical characteristics of Harris corner detector. In *Proc. SSP*, pages 571–575, 2007.

[6] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *IJCV*, 37(2):151–172, 2000.

[7] R.M. Steele and C. Jaynes. Feature uncertainty arising from covariant image noise. In *Proc. CVPR*, volume 1, 2005.