# Epipolar Constraints for Multiscale Matching

Bill Triggs
Bill.Triggs@imag.fr
Pashmina Bendale
pb397@cam.ac.uk

Laboratoire Jean Kuntzmann
BP 53, 38041 Grenoble Cedex 9, France
Signal Processing Laboratory, Department of Engineering
University of Cambridge, Cambridge CB2 1PZ, UK

Many recent keypoint detectors [3, 4] associate a local scale (for multiscale detectors) or even a full affine frame (for affine-invariant detectors) to each detected keypoint. Conventional epipolar geometry [2] only constrains the relative positions of corresponding keypoints, not their relative scales. We present an enhanced epipolar constraint that exploits both positions and scales, thus making correspondence search 2-4 times more accurate in practice.

The method works as follows:

- encode multiscale keypoints as image ellipses;
- invoke the 'Kruppa constraints' that link corresponding ellipses [2];
- project to the "epipolar pencil" (the 1-D family of epipolar lines) to get reduced constraints linking 1-D quadratic forms on the pencil;
- enforce a scale-sensitive (angular position, angular width) error model by a well-chosen algebraic transformation of this representation.

Figure 1 illustrates the projection process. The $2\times3$ matrices $\mathbf{B},\mathbf{B}'$ generating the projections onto the epipolar pencil are extracted from the singular value decomposition of the fundamental matrix [2] via $\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}^\top = \mathbf{F} = \mathbf{B}^\top \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \mathbf{B}'$, whence $\mathbf{B} \equiv (\mathbf{u}_2, -\mathbf{u}_1)^\top$ and $\mathbf{B}' \equiv (\mathbf{S}_{11}\mathbf{v}_1, \mathbf{S}_{22}\mathbf{v}_2)^\top$. Possibly-corresponding pairs of multiscale keypoints are represented as image ellipses – $3\times3$ symmetric dual-form conic matrices $\mathbf{q}$, $\mathbf{q}'$ – and projected onto the epipolar pencil by $\mathbf{B}\mathbf{q}\mathbf{B}^\top$ and $\mathbf{B}'\mathbf{q}'\mathbf{B}'^\top$. The reduced Kruppa constraints say that if the ellipses correspond, these two symmetric $2\times2$ matrices agree up to scale. Algebraically, this turns out to provide a strong constraint on the correspondence of keypoint centres, but only a weak second order one on the correspondence of their scales. To work around this, we algebraically transform the constraint to the following (keypoint position, keypoint scale) "normalized distance" model:

$$d_{\bar\theta} \equiv \frac{\sin^2 2(\bar\theta-\bar\theta')}{\sin^2\delta\theta+\sin^2\delta\theta'} = \frac{(pq'-qp')^2}{1-(r+r')/2} \tag{1}$$

$$d_{\delta\theta} \equiv \left(\frac{\sin\delta\theta}{\sin\delta\theta'}\right)^k + \left(\frac{\sin\delta\theta'}{\sin\delta\theta}\right)^k - 2 = \left(\frac{1-r}{1-r'}\right)^{k/2} + \left(\frac{1-r'}{1-r}\right)^{k/2} - 2 \tag{2}$$

Here, $(p,q,r)$ and $(p',q',r')$ (with scaling to $p^2+q^2=1$) linearly encode the 3 independent entries of the matrices $\mathbf{B}\mathbf{q}\mathbf{B}^\top$ and $\mathbf{B}'\mathbf{q}'\mathbf{B}'^\top$, and $(\bar\theta,\delta\theta)$ and $(\bar\theta',\delta\theta')$ are corresponding mean angle, angular width representations (with angles measured in projective coordinates). The above forms have appropriate small and large angle limits and close relationships to standard statistical error models. The formulae above assume standard full-epipolar-line matching but the method also handles signed (epipolar half-line) correspondance and omnidirectional images via a twofold unwrapping process based on oriented projective geometry.

The final method is very simple to use – perhaps even simpler than standard epipolar line search – and it gives good results on both synthetic images (see fig. 2) and real images (see fig. 3). In both cases, incorporating the additional scale constraint into the epipolar matching process cuts the number of false positive matches by a factor of 2–4 over a wide range of camera geometries and imaging conditions.

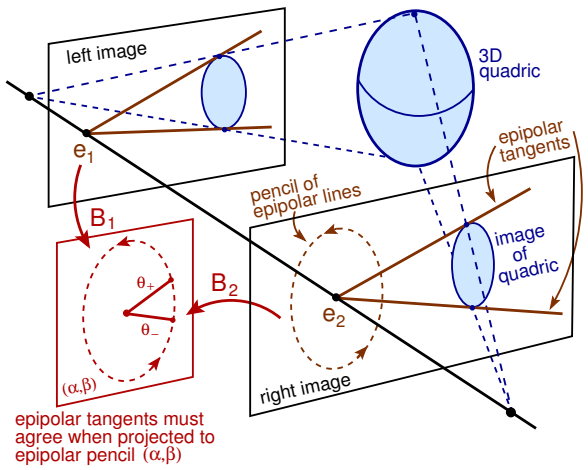


Figure 1: The projection of a 3D quadric to two image conics, evaluation of the pair of epipolar lines tangent to each conic, then further projection of the conics and their epipolar lines onto the epipolar pencil via $2\times3$ epipolar projectors $\mathbf{B},\mathbf{B}'$. The reduced Kruppa constraints state that the two projections agree.

[1] P. Bendale, B. Triggs, and N. Kingsbury. Multiscale keypoint analysis based on complex wavelets. In *British Machine Vision Conference*, August 2010.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[3] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Computer Vision*, 60(2):91–110, 2004.

[4] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *Int. J. Computer Vision*, 60(1):63–86, 2004.

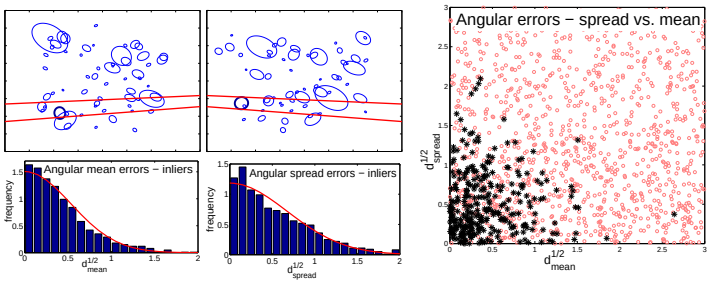


Figure 2: Experiments on synthetic data. *Top left:* left and right images of a scene containing random 3D ellipsoids, showing the epipolar lines tangent to a selected ellipse in the right image, and the corresponding lines in the left image almost tangent to the corresponding (noise perturbed) left ellipse. *Bottom left:* distributions of matching distance values for correct but noise perturbed correspondences, for (left) mean angle distance $\sqrt{d_{\bar\theta}}$, and (right) angular spread distance $\sqrt{d_{\delta\theta}}$. The $d$ penalties behave roughly like $\chi^2_1$ variables, *i.e.* the $\sqrt{d}$ plots resemble half-Gaussians (red curves). *Right:* scatter plot of $\sqrt{d_{\delta\theta}}$ versus $\sqrt{d_{\bar\theta}}$ values over a large dataset. The black '*'s are correct matches and the red 'o's are incorrect ones. Clearly, both the mean and the spread terms are useful for distinguishing inliers from outliers. The combined decision rule $d_{\bar\theta}/\langle d_{\bar\theta}\rangle + d_{\delta\theta}/\langle d_{\delta\theta}\rangle$ reduces false positives by a factor of about 2–2.5 relative to classical epipolar line thresholding (here, '$\langle-\rangle$' denotes empirical means over inliers).

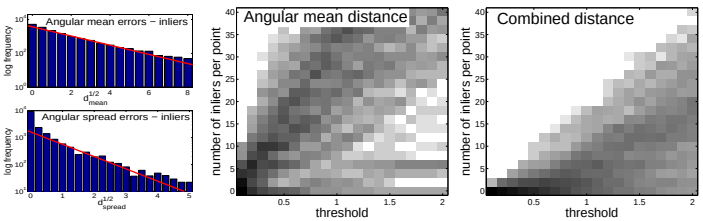


Figure 3: Experiments with SIFT keypoints on real images from [1]. *Left:* The distributions of $\sqrt{d_{\bar\theta}}$ and $\sqrt{d_{\delta\theta}}$ for corresponding features appear to be exponential with medians $m_{\bar\theta} \approx 1.0$ and $m_{\delta\theta} \approx 0.6$. *Middle:* Histogram of numbers of candidate matches satisfying the epipolar correspondence rule $\sqrt{d_{\bar\theta}}/m_{\bar\theta} < t$, for varying thresholds $t$. Darkness is proportional to log frequency. *Right:* The corresponding histogram for the combined decision rule $\sqrt{d_{\bar\theta}}/m_{\bar\theta} + \sqrt{d_{\delta\theta}}/m_{\delta\theta} < t$. The combined rule is about 4 times more selective, producing many fewer incorrect correspondences.