Multiscale Keypoint Analysis based on Complex Wavelets

Pashmina Bendale[1]
pb397@cam.ac.uk

Bill Triggs[2]
Bill.Triggs@imag.fr

Nick Kingsbury[1]
ngk@eng.cam.ac.uk

[1] Signal Processing Laboratory, Department of Engineering,
University of Cambridge, Cambridge CB2 1PZ, UK

[2] Laboratoire Jean Kuntzmann
BP 53, 38041 Grenoble Cedex 9 France

We describe a new multiscale keypoint detector and a set of local visual descriptors, both based on the efficient Dual-Tree Complex Wavelet Transform (DTCWT). The detector has properties and performance similar to multiscale Förstner-Harris [2] detectors. The descriptor provides efficient rotation-invariant matching, with performance similar to SIFT. Our work builds and significantly improves on an earlier DTCWT-based keypoint detector [1] and descriptor [3].

4S-DTCWT. The DTCWT uses $O(N)$ real arithmetic to produce 6 complex analytic, orientation-sensitive subbands at each level k, oriented at $(30d - 15)^\circ$ for $d = 1 \ldots 6$. It provides significantly better shift invariance and orientation selectivity than conventional real discrete wavelet transforms. Our method is based on a scale-space pyramid, 4S-DTCWT, that uses four interleaved scale-normalised DTCWT trees to provide denser scale sampling and hence reduce scale-related aliasing effects.

BTK keypoint detector. Our keypoint detector finds the maxima over position and scale of the novel cornerness measure

$$\widetilde{E}_k(x,y) \equiv \min_{\text{orientations } d \in \{1, \ldots, 6\}} |\widetilde{H}_k(x,y,d)|. \tag{1}$$

Here $\widetilde{H}_k(x,y,d)$ is the 4S-DTCWT wavelet coefficient at position (x,y), level k and orientation d. Subpixel positions are estimated by local $3 \times 3 \times 3$ quadratic fitting. The BTK detector detects Harris-like points, *i.e.* ones with strong 2D texture.

BTK descriptor. We refine the Polar Matching Matrix (**P**-matrix) descriptors of [3]. These are created by sampling DTCWT coefficients at a circle of 12 points around the keypoint position (6 orientations × 12 points) and appending DTCWT coeffients of two scales at the central position to create a 12×8 complex visual descriptor matrix. This is arranged so that rotations by multiples of 30° produce cyclic shifts within each column, thus allowing FFT to be used to rapidly match the orientations of pairs of descriptors. We build our **P**-matrices from 4S-DTCWT coefficients at the exact subpixel position and scale of the keypoint, using the wavelet level nearest to this scale.

These descriptors have similarities to other modern high-dimensional descriptors such as SIFT [4] and DAISY [6], but with a novel orientation-handling mechanism. Although not affine-invariant (with the current detector), like SIFT they tolerate small errors in keypoint positions and scales and small affine deformations relatively well.

Dataset and evaluation framework. We introduce a new dataset for the automatic evaluation of 3D viewpoint invariant visual correspondence, containing 4000 calibrated images of toy cars on a turntable. The associated evaluation framework uses three image epipolar geometry in a similar way to [5] – see fig. 1. The dataset and the evaluation software are publicly available[1].

We used this framework to quantitatively evaluate our methods, comparing them to selection of popular detectors and descriptors from the literature and to a previous DTCWT keypoint detector [1] ('FKA'). Detectors and descriptors were evaluated independently. Some of the results are illustrated in figures 2 and 3. Overall the performances of our multiscale DTCWT-based keypoint detector and descriptor are in line with those of other popular methods.

[1] J. Fauqueur, N. Kingsbury, and R. Anderson. Multiscale keypoint detection using the dual-tree complex wavelet transform. In *International Conf. Image Processing*, 2006.

[2] W. Förstner. A framework for low-level feature extraction. In *European Conf. Computer Vision*, pages 383–394, 1994.

[3] N. Kingsbury. Rotation-invariant local feature matching with complex wavelets. In *European Conf. Signal Processing*, 2006.

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

[5] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3D objects. *Int. J. Computer Vision*, 73(3):263–287, 2007.

[6] S. Winder and M. Brown. Learning local image descriptors. In *IEEE Conf. Computer Vision & Pattern Recognition*, 2007.
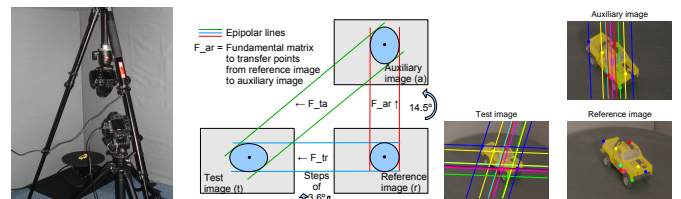
Figure 1: The set-up used for our dataset and experiments is similar to [5]. Left: the rig used to capture the dataset. Middle: The epipolar matching scheme. For each keypoint in the reference image, epipolar constraints and normalised colour cross-correlation are used to estimate the location of the corresponding keypoint in the auxiliary image. The intersection of the reference-image and auxiliary-image epipolar lines in the test image then gives us the predicted location of the corresponding keypoint in the test image (if any). Right: Some example images with epipolar lines.
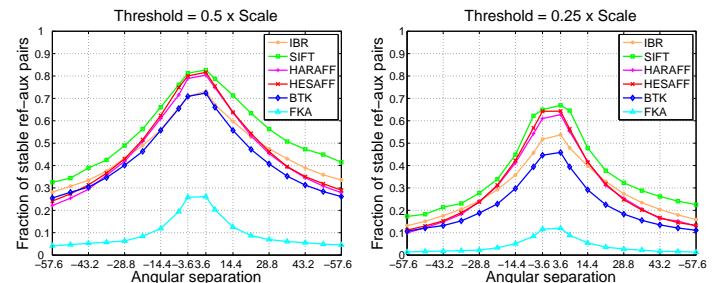

Figure 2: Repeatability of various keypoint detectors under changes in viewpoint. We plot the fraction of reference-auxiliary pairs that have a test-image keypoint of the estimated scale (radius) at the estimated location, for a range of angular separations and for two acceptance thresholds. A gradually falling curve indicates good tolerance to changes of viewpoint. Note that the tolerance degrades as the constraint on localisation error is tightened. Our BTK DTCWT detector improves greatly on the existing FKA one owing to its better position and scale estimation. Its performance is now in line with other established detectors. All of the detectors here find about 100 ± 5 reference-auxiliary pairs per image tested.
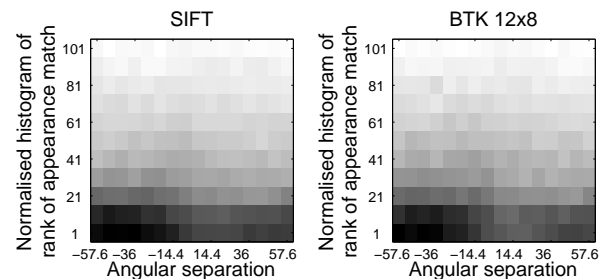

Figure 3: Descriptor repeatability of under changes of viewpoint. We plot normalized histograms of the rank of the true-correspondence keypoint among all of the keypoints in the image, where rank is measured by inter-descriptor distance to the target descriptor. These are plotted as a function of the inter-image angle, for SIFT and BTK descriptors over SIFT keypoints. Darker colors indicate higher frequencies. The figure shows that (unlike keypoint detectors) increasing angular separation has relatively little effect on these descriptors.

[1] http://www-sigproc.eng.cam.ac.uk/imu/