Multiscale Keypoint Analysis based on Complex Wavelets

Pashmina Bendale, William Triggs, Nick Kingsbury

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

Abstract

We describe a new multiscale keypoint detector and a set of local visual descriptors, both based on the efficient Dual-Tree Complex Wavelet Transform. The detector has properties and performance similar to multiscale Förstner-Harris detectors. The descriptor provides efficient rotation-invariant matching. We evaluate the method, comparing it to a previous wavelet based approach and to several conventional detectors and descriptors on a new dataset designed for the automatic evaluation of 3D viewpoint invariance. The dataset contains over 4000 images of toy cars on a turntable under accurately calibrated conditions. Both it and the evaluation software are publicly available. Overall the method gives performance competitive with existing Harris-like detectors.

1 Introduction

We present an approach to multiscale keypoint matching based on the critically sampled Dual-Tree Complex Wavelet pyramid. Although wavelets have proven very successful for image compression, image coding, denoising and deconvolution, there has been little work on using them for local descriptor based image matching. Similarly – although some phase-based approaches do exist, *e.g.* [1] – most of the existing work on multiscale keypoint detection is based on conventional real representations such as Difference of Gaussian decompositions [8, 10], not on wavelets or complex representations.

This paper describes a new multiscale keypoint detector and a set of local visual descriptors, both based on the efficient Dual-Tree Complex Wavelet Transform (DTCWT) [13]. We describe the method and evaluate it, comparing it to a previous wavelet based approach and to several conventional detectors and descriptors on a new image dataset designed for the automatic evaluation of viewpoint invariant local feature methods.

We summarise the existing wavelet based approaches in §2, develop our new 'BTK' multiscale keypoint detector based on the DTCWT in §3 and detail our DTCWT based local descriptor in §4. We introduce our new dataset in §5, describe the evaluation method and present results in §6, and summarise in §7.

2 Background

Dual-Tree Complex Wavelet Transform. The DTCWT of a 1D signal [13] uses a carefully designed dyadic tree of Hilbert Transform pair filters to compute the real and imaginary components of a complex analytic wavelet decomposition using only efficient real arithmetic. In the 2D image setting, efficient paired DTCWT trees can be designed [13] to output six directionally sensitive, approximately analytic subbands oriented at angles $(30d-15)^\circ$ for $d = 1\ldots6$. The rotational symmetry of the DTCWT can be further improved by adding a bandpass filter in each direction and applying a phase correction to make the coefficients conjugate symmetric [7]. This provides significantly better shift invariance and orientation selectivity than conventional real discrete wavelet transforms (DWTs) at lower computational cost than a comparable steerable filter [14].

FKA Keypoint Detector. Our work builds on, and significantly improves, an earlier DTCWT based keypoint detector [3]. This earlier version detects keypoints at the maxima of the "accumap" – an accumulated map of responses across scale and orientation, $\sum_{\text{tree levels } k} E_k(x,y)$, where $E_k(x,y) \equiv \prod_{\text{orientations } d=1..6} |H_k(x,y,d)|^{1/4}$ and $H_k(x,y,d)$ is the complex DTCWT coefficient at level k, subband (orientation) d and location (x,y). The moduli of the wavelets characterise the oriented gradient energy at the given position and scale, so taking their product over orientations gives a response reminiscent of a Harris (determinant of oriented energy tensor) detector. Summing these responses over all tree levels provides a degree of scale invariance. Given the (x,y) position of an accumap maximum, the scale of the corresponding keypoint is estimated by searching for the first radial distance at which the negative gradient of the accumap has a strong maximum, as measured in the sum of gradients along 8 radial directions spaced by 45°.

Although this detector works to some extent, in practice (see §6) we find that it tends to produce false detections on strong edges due to aliasing and imperfect alignment of filters with edge orientations, and that its scale estimates are too inaccurate to support reliable visual descriptor based matching of the resulting keypoints. We therefore developed an improved DTCWT keypoint detector, BTK, which we present in the next section.

3 BTK Keypoint Detector

Our basic goal was to develop a DTCWT based detector that does not fire inappropriately on edges and that provides accurate subpixel keypoint position and scale estimates that transform appropriately under small translations and dilations (spatial rescalings) of the input signal. Regarding the second point, for a minimally sampled transform, the DTCWT already provides an exceptionally well controlled response to translations owing to its complex analytic design, so no improvement is needed there. Unfortunately, the same can not be said of scale resilience: although complete as a representation, dyadic (power-of-two) wavelets are too coarsely sampled in scale to prevent substantial aliasing of energy between levels under small dilations of the input signal.

To remedy this our method uses several (in practice four) DTCWT trees rather than just one, interleaving them in scale to provide denser scale sampling. Ideally they would be spaced at scale intervals of $2^{-1/4}$, but for implementation reasons it is easier to space them at scales of $[1, \frac{7}{8}, \frac{6}{8}, \frac{5}{8}]$ and experimentally we find that four trees with these spacings suffice for good results. The method rescales the input image by these amounts using bilinear interpolation and then evaluates a separate DTCWT tree over each replica (as usual, decimating

by factors of 2 and padding the result to an even size as needed for the calculation of further levels). If the basic DTCWT of the image has K levels we evaluate $K-1$ levels in each of the three new trees, thus producing a pyramid with $4K-3$ levels in all.

For invertibility the original DTCWT uses energy-preserving wavelet normalisation, whereas image resampling typically aims to preserve the range of gray-level values of the local signal, not its energy. We adopt gray-level based normalisation to ensure that keypoint responses, detection thresholds, *etc*., are independent of scale. Hence we scale down the level-k DTCWT wavelet coefficients by 2^{-k} (*i.e*. $2^{-k/2}$ for each of the 2 dimensions of the image). We will refer to the resulting scale-normalised 4-tree DTCWT as the **4S-DTCWT**[1], using $\widetilde{H}_k(x,y,d)$ to denote the corresponding wavelet coefficients. 4S-DTCWT remains computationally efficient, requiring about $1+(\frac{7}{8})^2+(\frac{6}{8})^2+(\frac{5}{8})^2 \approx 2.7$ times the computation required by the native DTCWT, plus the cost of the initial image resampling.

Secondly, although we tested a Harris-like single-level keypoint strength function of the geometric-mean form $\widetilde{E}_k(x,y) = \prod_{d=1}^{6} |\widetilde{H}_k(x,y,d)|^{1/6}$, we prefer to use a function of the form

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

This is somewhat analogous to using the minimum eigenvalue of the oriented energy tensor (Förstner detector [4]) rather than its determinant (Harris detector [6]), as the discrete minimum over orientations spaced by 30° is similar to the continuous minimum over all orientations[2]. Overall the performance of the two approaches is similar but the one adopted gives slightly better rejection of false responses along edges, and is also faster to evaluate. Note that in both cases our detector is Harris-like in the sense that it finds regions (keypoints) that have strong oriented edge energy at multiple orientations, and thus sharply defined local autocorrelation functions. This is in contrast to methods (Difference or Laplacian of Gaussian, Hessian, SIFT, *etc*.) that key on "blob"-like structures but not on arbitrary strong 2D textures. Blobs occur less frequently than 2D textures, but their clearly defined positions and scales make them particularly suitable for descriptor based matching. In practice both kinds of detectors are useful and it would be interesting to develop a DTCWT based blob detector.

Thirdly, rather than accumulating the responses (1) across scales like [3], we build a pyramid from the single-scale responses, search this for local maxima over position and scale, and interpolate these to subpixel accuracy in position and scale. Specifically, at each level of the pyramid we find 2D local maxima over 3×3 local patches of the response function, extract 3×3×3 scale space patches around these by importing the 3×3 patches nearest to the location of the 2D maximum at the scales immediately above and below it, discard 2D maxima that are not local maxima over their full 3×3×3 patch, use least squares to fit a local quadratic to the function samples on the patch, and use the peak of the quadratic as the position, scale, and value of the maximum. The quadratic fitting is done in 'expanding' local coordinates around the 2D maximum, using log scale for the scale coordinate and using (x,y) coordinates whose origin is the image location of the 2D maximum sample, transferred to its equivalent subpixel location at each of the three scales, and whose available (pyramid) samples are separated by $|\Delta x|, |\Delta y| = 1$ at each of the three scales[3]. Note that samples at

[1] This is not a surjective wavelet transform, just a method for computing a scale-space pyramid.

[2] We also attempted to estimate the angular minimum embodied in (1) more accurately using inter-orientation interpolation and using angular Fourier expansion, but this did not improve the resulting detector. It seems that 30° sampling is not fine enough to make interpolation over the necessary 3–4 adjacent samples a reliable predictor of the true response at intermediate orientations.

[3] This rather strange local coordinate system is the best one to use for many local scale space computations. To the extent possible, it mediates between the fact that responses typically broaden in proportion to sample spacing as

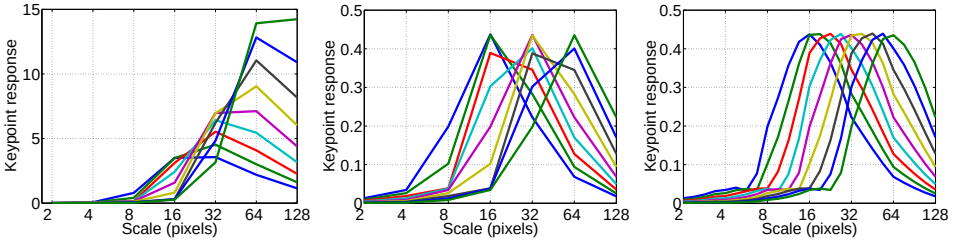

Figure 1: Scale responses to a set of images containing 2D Gaussian blobs with fixed amplitude and varying widths ($\sigma = 4$ to 16 in steps of $2^{1/4}$). Each curve shows the response for a single image, plotted as a function of pyramid scale at the (x,y) peaks of the single scale responses (1). Left: standard DTCWT without scale normalisation (as used in [3]). Middle: standard DTCWT with 2^{-k} scale normalisation. Right: 4S-DTCWT, with both scale normalisation and denser scale sampling. Notice the extent to which 4S-DTCWT provides both better-defined peaks and more consistent response amplitudes over scale. This leads directly to more consistent scale estimates.

different scales are offset relative to one another and that in our "8-7-6-5" 4-tree pyramid adjacent levels are not uniformly spaced in scale. The least squares fit uses exact position and scale values for each sample point to compensate for this. Although such $3\times3\times3$ fitting is commonly used for subpixel peak finding, its results do depend significantly on the sample chosen as centre (*e.g.* if several adjacent samples have essentially the same keypoint response). We also tested a subpixel peak finder based on a Gaussian mean-shift like procedure in the same coordinate system, formulated in such a way that it handled the resulting uneven sampling appropriately. However we found that (despite its elegance) this method required samples from too many scales, leading to large computational overheads and frequent losses of detections owing to the mean shift drifting to peaks at larger or smaller scales. Hence, in the end we preferred the simple $3\times3\times3$ quadratic fit method.

Fig. 1 illustrates the extent to which 4S-DTCWT improves the scale invariance of the response function (1), by showing the responses arising from a set of 2D Gaussian shaped blobs of fixed amplitude and slowly changing spatial scale.

4 BTK Local Descriptor

To complement our detector we use rotation-invariant DTCWT-based 'Polar Matching Matrices' as the local visual descriptor [7] for our keypoints. We briefly present these descriptors, then show how we adapt them for use with BTK keypoints.

Polar matching matrix (**P**-matrix) descriptors are created from DTCWT coefficients as follows [7]. At a designated DTCWT level and sampling radius, a circle of 12 points spaced 30° apart is placed around the central point (keypoint), and for each DTCWT orientation, its complex DTCWT coefficient is evaluated at each point, using spatial interpolation in the DTCWT map as necessary[4]. At each point there are 6 independent orientations spaced 30° apart (and their complex conjugate pairs spaced 180° away from these). The resulting coefficients are arranged in a 12×6 complex matrix whose column c contains the coefficients

the scale changes, and the need to align corresponding image positions across different scales.

[4] Such interpolation is reliable owing to the band-limited nature of the rotationally symmetric DTCWT.

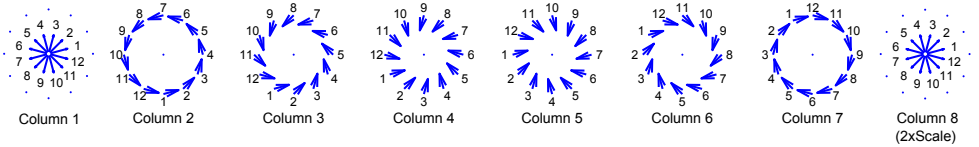

Figure 2: The construction of our 12×8 **P**-matrix descriptor. The arrangement is the same as in [7], but we base the descriptor on the 4S-DTCWT. The small numbers (and the orientations of the arrows) denote the subband selected at each sampling location. Each column is composed of a set of rotationally symmetric samples. The figure is taken from [7].

whose orientation relative to the tangent to the sampling circle at the sample position is $(30c-15)^\circ$. Rotations of the image by multiples of 30° thus produce cyclic shifts within each column of the matrix, *i.e.* simple phase changes of the FFT of the column. This property allows efficient rotation-invariant descriptor comparison and estimation of relative rotations between descriptors. To produce a complete **P**-matrix descriptor, matrices from several circles with different radii and/or wavelet scales (tree levels) can be appended, and additional columns can be included based on the coefficients of the 12 orientations (6 conjugate pairs) at the central point at a given level. One of the most common arrangements [7] is a spatially-compact local descriptor whose 12×8 matrix contains the coefficients from both the central point and the circle with radius one sample spacing at the given level, together with those from the central point at the next level up (wavelets 2× coarser) – see fig. 2. For illumination invariance, the total energy in each **P**-matrix is normalised to one, so matching them produces a correlation score in the range $[-1,1]$.

To use these descriptors with our detector, we need to adapt them to subpixel position and scale estimates. We use the 4S-DTCWT for the descriptors as well as the detector so raw coefficients are available at step sizes of around $2^{1/4}$ in scale and at integer locations at these scales. Given a keypoint, we lay out a circle of subpixel sample points corresponding to its exact subpixel location and scale (the circle having unit radius at this scale). We then build the 12×8 **P**-matrix descriptor by taking wavelet coefficients from the discrete 4S-DTCWT level whose scale is closest to the keypoint scale and using subpixel interpolation to estimate the wavelet responses at the designated sample points. (For the 8^{th} column, we use the level 2× higher). The descriptor is thus evaluated at sample points corresponding to its exact subpixel position and scale, but using wavelet coefficients from the nearest discrete scale. We will refer to these 12×8 4S-DTCWT **P**-matrix descriptors as 'BTK descriptors'.

The BTK descriptor has some similarities to other modern descriptors such as SIFT, DAISY [16], *etc*. It is based on oriented energies at several spatial positions, it incorporates multiple scales and good illumination invariance, and it has an effective (but original) mechanism for handling the overall orientation degree of freedom. Although (with the current detector) it is not affine-invariant, like SIFT it tolerates small errors in keypoint positions and scales and small affine deformations relatively well. Fig. 3 illustrates that under increasing affine deformations, BTK descriptor based matching scores show degradations similar to those for SIFT, but perhaps slightly less rapid at small deformations.

5 3D Dataset and Test Framework

Previous evaluations of keypoint detectors and descriptors have tended to focus on planar scenes, relying on homographies to generate the ground truth [9, 12]. Evaluations on 3D

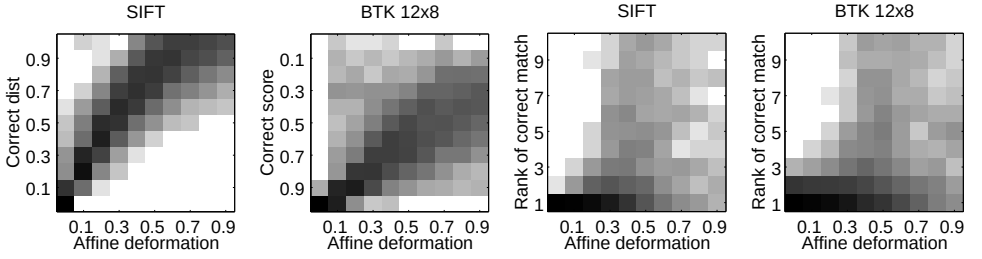

Figure 3: Descriptor mismatch under affine deformations. SIFT and BTK descriptors are computed and matched over identical (Difference-of-Gaussian) keypoints for increasing affine distortion (vertical shear) of a Graffiti image from the Oxford dataset [10]. The leftmost two plots show histograms of the resulting SIFT distances (0 is best) and DTCWT correlation matching scores (1 is best). The rightmost two plots show histograms of the rank of the correct match using these respective distance metrics for ranking. In all cases, darker colours indicate higher bin counts with darkness proportional to log(count). Note that the distribution of ranks for BTK is similar to that for SIFT.

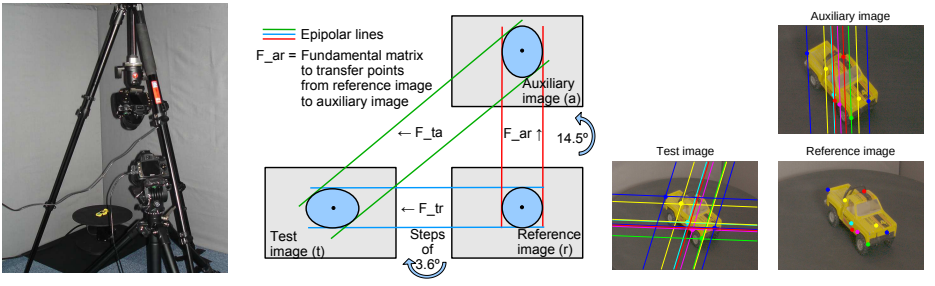

Figure 4: The set-up used for our dataset and experiments is similar to [11]. Left: the rig used to capture the dataset. Middle: The epipolar matching scheme. For each keypoint in the reference image, epipolar constraints and normalised colour cross-correlation are used to estimate the location of the 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.

scenes include [5] (using trifocal tensors on natural scenes) and [11] (using calibrated images from a turntable). In order to provide an accurate evaluation of methods on real scenes, we wanted to use a method similar to [11]. Unfortunately, although these authors provide a robust framework, they have not made their complete ground truth, calibration and test software publicly available. Moreover, the objects typically only occupy a small portion of their images and there are typically many detections on the image backgrounds, which are neither plain nor realistic. For this reason we decided to create a new dataset containing 4000 calibrated images from 2 cameras of 40 different toy vehicles on a turntable with a plain grey background. There are 51 images per car per camera spanning 180°. The images in both PNG and raw (NEF) format, the calibration images and camera parameters including lens distortions, and the MATLAB test scripts are all freely available for download[5]. We used the publicly available DLR CalDe and CalLab toolbox [15] for calibration and DCRAW [2]

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

for decrypting the NEF (Nikon raw format) files. Although we did not use them in this study, convex polygonal boundaries of the cars in both cameras are also available for 15 sequences. In this dataset, there are typically around 200–500 keypoint detections on the object and only a few (~ 10) weak detections on the background, so the overall statistics are not greatly affected by background points.

Geometrically, both cameras are at approximately the same distance from the centre of the turntable and looking directly at it. One ("auxiliary") is above and somewhat in front of the other ("reference/test"), the difference in elevation being about 15°. This arrangement produces near vertical epipolar lines between corresponding reference and auxiliary images, near horizontal ones between adjacent reference/test images, and similar image scales in all images (which simplifies the evaluation of keypoint scale estimates). For the evaluations, for each "reference/test" image in turn ("reference"), we use epipolar geometry against its corresponding "auxiliary" image to find possible matches, then use 3-image epipolar geometry to evaluate whether a corresponding point was found in the designated "test" image (the "reference/test" one at a given angular separation from the current "reference"). Fig. 4 illustrates the set up.

6 Experiments on our 3D Dataset

We tested the BTK keypoint detector against other methods including the original FKA DTCWT based detector [3], the SIFT Difference of Gaussian detector [8], and the Intensity Based Region (IBR), Harris-Affine (HAR-AFF) and Hessian-Affine (HES-AFF) detectors [10]. The most comparable previous evaluations are [11] and [5]. To the extent possible, we separated the evaluation of the keypoint detectors from that of descriptors.

In contrast to previous 3D studies, we base the thresholds used for inlier decisions on the scale of the reference keypoint. We also ran comparable tests using fixed thresholds – *e.g.* 5 pixels from the epipolar intersection, irrespective of keypoint scale. This simply translated all of the curves vertically by an amount depending on the threshold, with no noticeable change in shape or relative ranking. The fixed-threshold scheme is biased towards fine scale keypoints in the sense that it allows greater relative localisation errors for these than for coarse scale ones. Also, when local descriptors are used, they are computed on scale-normalised image patches and localisation errors should be measured in terms of their effect on the descriptors, *i.e.* as a function the scale of the keypoint.

We find that gamma compression of the image – $\mathbf{I} \leftarrow [\mathbf{I} + c]^{\gamma}$ with $c \sim$ 20–30 and $\gamma \sim$ 0.3–0.5 – improves the performance of all of the keypoint detectors except SIFT (which has built-in illumination invariance). We use this correction in all of the tests.

Detector repeatability. Fig. 5 summarises the results of an evaluation of the repeatability of various keypoint detectors under changes in viewpoint, using scale-dependent thresholds. The test range is from $-57.6°$ to $57.6°$ in steps of $3.6°$ and excluding $0°$. Only reference and auxiliary images from the central $60°$ of the $180°$ sequences are used so that each ref-aux pair will have the full $\pm 57.6°$ range of test viewpoints. Markers are shown at steps of $7.2°$ with extra points at $3.6°$ for better visibility near zero. The thresholds are set so that on average, each detector finds 100 ± 5 ref-aux keypoint pairs per image tested. For each viewing angle, we measure the probability for the detector to redetect a keypoint at the corresponding physical location and scale on the object. To do this, we locate the auxiliary-image point corresponding to a given reference point by searching for the best match using normalised colour cross correlation among the keypoints within a certain scale-independent

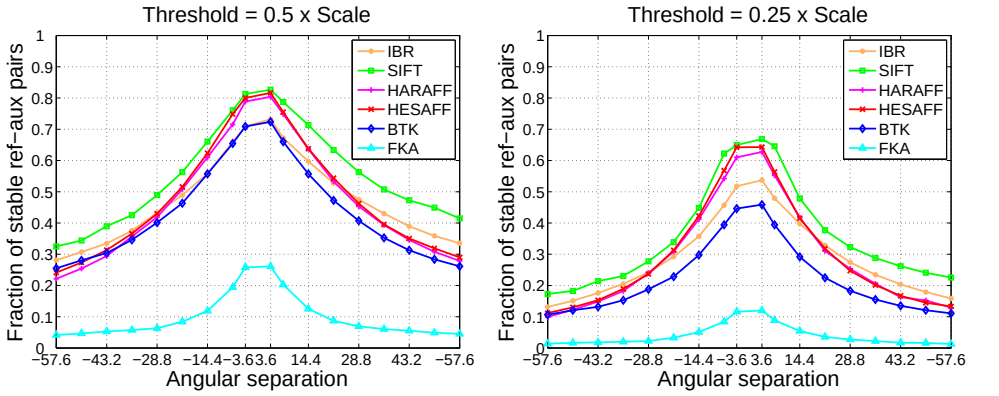

Figure 5: Repeatability of various keypoint detectors under changes in viewpoint. We plot the fraction of reference-auxiliary pairs that have a test-image keypoint of the estimated scale (radius) at the estimated location, for a range of angular separations and for two acceptance thresholds. A gradually falling curve indicates good tolerance to changes of viewpoint, but the curves become more peaked as the constraint on localisation error is tightened. Our BTK detector improves significantly on FKA owing to its better position and scale estimation, and its performance is now in line with that of the other established detectors.

distance of the reference point's epipolar line in the auxiliary image. The selected ref-aux pairs are used to predict keypoint positions in the test image by epipolar line intersection, and if a keypoint is detected within a certain distance of this position – 0.5× the scale of the reference keypoint for Fig. 5 (left) and 0.25× the scale for Fig. 5 (right) – we consider it to be a successful detection. Unsurprisingly, Harris-Affine, Hessian-Affine and SIFT detectors show very similar performance: all these detectors are based on Difference-of-Gaussian pyramids. Note the extent to which the BTK detector improves on the FKA one, owing to its better position and scale estimation.

Descriptor repeatability. Fig. 6 summarises an evaluation of repeatability under changes in viewpoint, for SIFT and BTK descriptors over SIFT keypoints. Similar results are obtained for BTK keypoints. For each viewing angle, we measure the likelihood of a good matching rank (and hence typically a good matching score) for the descriptors of points that are known to be geometric inliers. For this, we detect keypoints in all of the images, compute their descriptors, and find all of the geometric matches for the given viewing angle. For each such match we compare its reference descriptor with the descriptors of all points in the given test image and find the rank of the geometric match in that list. If the descriptor similarity remains roughly constant as the viewing angle increases, the distribution of ranks should also remain roughly constant because there are about the same overall number of detections in each image, whereas if the similarity degrades, the rank should increase with angle. To quantify this, for each descriptor and keypoint type, we plot the histogram of its ranks for each angle[6]. We see that – conditioned on the associated point being detected at all – the descriptor similarity is indeed roughly constant with angle[7].

Efficiency. On a 1536×1024 image with 389 keypoints, our current MATLAB implementa-

[6]For display purposes, each histogram (column) is normalised to sum 1. Although the overall number of detections remains roughly constant with angle, the number of detections that are geometric inliers decreases with increasing angle as we saw above (*i.e.* on average, old points that are no longer detected are replaced with new and

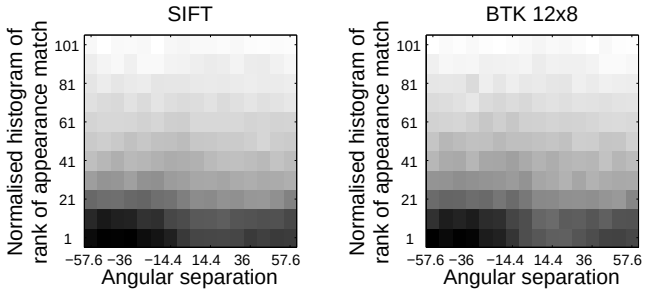

Figure 6: Rank histograms illustrating the repeatability of (left) SIFT descriptors and (right) BTK descriptors under changes of viewpoint, for SIFT keypoints. Darker colours indicate larger bin counts. We used a threshold of $0.5\times$ the scale of reference keypoint to search for geometric inliers. See the text for a full explanation.

tion takes 11.5 seconds: 6.3 to evaluate the 4S-DTCWT pyramid; 4 to detect keypoints; and 1 to compute their descriptors. A C implementation would probably be at least twice as fast. In comparison, Lowe's C implementation of SIFT [8] takes 5.5 seconds on the same image for 329 SIFT keypoints. In both cases the runtime is roughly linear in the total number of pixels in the image.

7 Summary

We have shown that the 4-Scale Dual-Tree Complex Wavelet Transform (4S-DTCWT) provides an efficient framework for keypoint detection and description based on minimally sampled wavelet responses. The resulting multiscale keypoint detector, BTK, has comparable performance to the most popular existing detectors. It detects a variety of features including blobs, corners and high-curvature points. This complements other region-based detectors like MSER, EBR and IBR. The associated rotation-invariant BTK descriptors provide an efficient local appearance description whose performance is comparable to that of SIFT descriptors, using a matching technique that does not require the estimation of dominant orientations during keypoint detection. Overall, the 4S-DTCWT approach to multiscale analysis makes a useful addition to the available keypoint detectors and descriptors.

Acknowledgements. P.B. would like to thank Dr. Jonathan Cameron for his help in making the dataset and the Gates Cambridge Trust for its financial support. Both P.B. and B.T. would like to thank the European network PASCAL2 for its travel support.

References

[1] G. Carneiro and A. Jepson. Flexible spatial configuration of local image features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2089–2104, 2007.

different ones). Hence different inlier histograms have similar distributions but very different numbers of counts.

[7]The epipolar geometry of our 3-image setup is weaker at large angles owing to shallower epipolar lines intersection angles, thus producing a wider search region for matches at these angles. This in turn slightly decreases the observed ranks at large angles, as more candidates are tested. This effect is purely geometric and not due to the descriptor performance.

[2] D. Coffin. DCRAW: Decoding raw digital photos in linux. Available at http://cybercom.net/~dcoffin/dcraw, 2008.

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

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

[5] F. Fraundorfer and H. Bischof. A novel performance evaluation method of local detectors on non-planar scenes. In *IEEE Conf. Computer Vision & Pattern Recognition - Workshops*, volume 03, page 33, 2005.

[6] C. Harris and M. J. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–152, 1988.

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

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

[9] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[10] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. J. Computer Vision*, 65(1-2):43–72, 2005.

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

[12] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.

[13] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, 2005.

[14] E. Simoncelli and W. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *International Conf. Image Processing*, 1995.

[15] K. Strobl, W. Sepp, S. Fuchs, C. Paredes, and K. Arbter. DLR CalDe and DLR CalLab. Institute of Robotics and Mechatronics, German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Available at http://www.robotic.dlr.de/callab/.

[16] S. Winder and M. Brown. Learning local image descriptors. In *IEEE Conf. Computer Vision & Pattern Recognition*, 2007.