PHASE INVARIANT KEYPOINT DETECTION

*Anil Anthony Bharath**

Department of Bioengineering
Imperial College London, UK
a.bharath@imperial.ac.uk

Nick Kingsbury

Department of Engineering
Trumpington Street, Cambridge, UK
ngk@eng.cam.ac.uk

Abstract

This paper introduces extensions to the keypoint detection paper [1]. Keypoints are generated by finding local peaks in accumulated, interpolated maps of the product of magnitudes of directional complex filter responses, as in earlier work. Gradient vector fields derived from these maps are used for keypoint scale characterisation, but this is now performed so as to remove the directionality of gradient field sampling, thereby improving the stability of scale estimates. A new class of keypoints is also introduced: the Circular Measure (CM) keypoints, which are used to augment the locations found by the filter magnitude product (FMP) keypoints. These new keypoint types generate a higher proportion of keypoints in the interiors of objects whilst simultaneously providing approximate object scale information, and appear appropriate for directing the attention of a vision system to the interiors of well-defined regions.

1. INTRODUCTION

Keypoints have been widely and successfully used to identify locations in an image which contain distinctive visual structure. This distinctiveness leads to their application in determining image patch correspondence, and hence for estimating transformation parameters between multiple views acquired in computer vision systems. Groupings of patches around distinctive feature points may also be used for object recognition, as demonstrated by Lowe's popular SIFT technique [2].

With the possible exception of scale determination, the SIFT approach may be split into two phases. The first is the identification of keypoints, and the second the characterisation of the regions around those keypoints by generating descriptors of each keypoint neighbourhood. We address only the first of these in this paper.

The detection of keypoints in SIFT is performed by a pyramidal scheme, in which successively blurred versions of the image are subtracted from each other in order to yield a hierarchy of "Difference-of-Gaussian" images. Due to the separability and semi-group property of Gaussian kernels, this is a relatively efficient operation, made more so by a standard dyadic multirate scheme. Lowe's implementation also constructs the intermediate scales of Gaussian smoothing (up to 3) for each level of decimation. The success of the keypoint approach has motivated substantial work into investigating potential improvements into the speed, precision, stability and reproducibility of keypoint detection. The approach adopted in this paper is to construct keypoints based upon the strong numerical properties of the DTCWT.

In this paper, we apply the Q-Shift version of the Dual Tree Complex Wavelet Transform (DTCWT) to identify keypoints of potential interest in the image. In addition to inheriting the benefits of near shift invariance in keypoint detection, an additional advantage of using complex wavelets is that the keypoints may be made robust to phase variations due to changes in viewing conditions. This may lend itself to better matching performance if contrast parameters change dramatically, including contrast inversion, and also to comparing line drawings or sketches with images of solid objects. We also propose an extension to the use of complex filter magnitude products, by creating a different class of keypoints that are sensitive to the interiors of objects. We term these Circular Measure keypoints, or CM keypoints.

*This work was partially completed during collaborative visits funded by by UKRC Grant GR 87642/02

2. IMPLEMENTATION DETAILS

2.1. Q-Shift DTCWT

The Dual Tree Complex Wavelet Transform (DTCWT) is described in detail elsewhere [3]. It has been applied to a variety of problems in image processing, where its near shift invariance, low redundancy, and strong directional selectivity prove useful. Moreover the modified DTCWT, recently introduced in [4], and employing frequency-shifted responses for the 45° and 135° channels, provides approximate rotational symmetry for the 6 directions of filtering at each scale. We do not discuss the implementation of the modified DTCWT here. For more details, the reader is directed to [4].

2.2. Filter Magnitude Products

The complex filter outputs, $f_k^{(\ell)}$, within one level, ℓ, of a pyramid may be used for efficient scale-selective corner detection [5]. Fauqueur, [1] proposed using filter magnitude products with a 1/6 power scaling as

$$P^{(\ell)} = \alpha^{\ell} \left(\prod_{k=1}^{6} |f_k^{(\ell)}| \right)^{1/6} \quad (1)$$

as feature point detectors, and suggested a means of combining FMP's across levels, by using a 2D interpolation process. α provides a tunable gain to the scale of response, allows corner, or interest point feature responses, to be made approximately scale invariant. See [5] for an alternative strategy. The first enhancement of this paper is a scheme for removing keypoints that are induced by jaggedness in images that are poorly bandlimited. This is typically found in images subjected to rescalings, rotations or general affine distortions in which inadequate interpolation has been used; strong directional responses at fine scales that are associated with jagged edges may have to be removed. Similar to Lowe and approaches adopted for scale-space medial axes [6], we compute the 2D Hessian matrix, and sort its eigenvalues. In accumulated FMP map regions that are peak-like, both eigenvalues will be large, negative, and close to each other in

value. Thus, the difference in eigenvalues of negative curvature provides information to discriminate between a peak and a ridge in this accumulator space. Dividing this difference by the largest eigenvalue gives a bounded measure that may be thresholded to reject FMP keypoints that lie on ridges, such as keypoints 7 and 8 in Figure (2).

2.3. Intrinsic Scale of FMP Keypoints

In order to characterise the intrinsic scale of the FMP keypoint over image locations, several approaches have been reported in the literature. Lindeberg [7] uses extrema over scales in normalised scale-space to optimally detect image structures at their intrinsic scale. Kadir and Brady [8] have shown that normalised maxima of entropy of low-level features in images can be used to detect salient image structures. The main problem with these approaches is that exhaustive filtering with kernels over a wide range of finely-sampled scales is computationally intensive and inefficient. Fast implementations of scale selection by searching for scale-space maxima has been proposed by [9, 10] using quadratic interpolation of the filter responses across scales. An analytical formulation for "steering" filter responses across the scale parameter is not so forthcoming because of the unbounded nature of the problem, i.e. filter scales can theoretically increase to infinity. In practice, the maximum filter scale is limited by the size of the image. Bharath [11] constructed exemplar vectors of the radial frequency response of filters across scales and used a Moore-Penrose generalised inverse to learn the steering coefficients through simple matrix algebra. Ng [12] modified this technique by learning the steering coefficients from radial response exemplars in the spatial domain and parameterised these coefficients as a function of scale by fitting polynomials.

The current approach is quite similar to that used in [1], in that the gradient field of the accumulated FMP response map is used. However, instead of extracting profiles through the FMP response gradient-field, vector field kernels that describe radially divergent fields from a candidate keypoint location are applied to the gradient field by a complex inner product. Each vector kernel, $M(\rho)\vec{U}(\phi)$ contains a radially varying weighting that peaks at a different radius. The kernels are constructed in a polar coordinate system, with origin at the keypoint, and the variation of the magnitude of the field, $M(\rho)$, is set by

$$M(\rho) = \frac{\beta^{n+1}}{\Gamma(n+1)}(\rho^n)e^{-\beta.\rho} \tag{2}$$

with $\beta = 4$ and with $n = \rho_p\beta$, where ρ_p is the radius at which peak radial sensitivity to gradient direction is desired. The angular characteristic for these kernels is defined by $\vec{U}(\phi)$, with

$$\vec{U}(\phi) = \cos(\phi)\vec{u}_x + \sin(\phi)\vec{u}_y \tag{3}$$

which yields a vector field outwardly diverging from the centre of the kernel.

The size of the kernels is set to 61×61, and 30 kernels are used, each with a different peak radial sensitivity. Because of the ability to represent vector fields using complex numbers, the kernels can be represented as unravelled complex arrays, so that all 30 kernels can be represented by one 30×3721 complex array. Note that kernels are not applied by convolution, but by a local complex inner product between patches unravelled into a 3721 length array. Thus, the computational load associated with convolution between images and large kernels is not significant unless the number of keypoints becomes excessively high.

Unraveled patches of the accumulated FMP space are multiplied by the matrix of complex kernels, stacked into columns, in order to generate smoothed profiles of radial gradient projections at 30 different radii; an example of a pair of such profiles is shown in Figure (1). Looking for local minima seems to be appropriate for characterising the scale of a coherent object. Consider Keypoint 12, corresponding to the nested circles of the test phantom introduced in [1] and illustrated in Figure (2). We compare this with the signature for Keypoint 13, which is generated by a single solid circle in the same figure. Note that there are two minima for Keypoint 12, but only one for Keypoint 13.

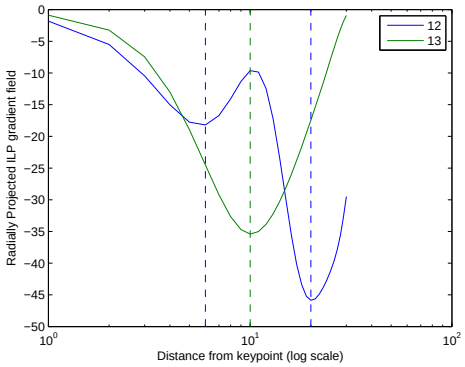

Fig. 1. Gradient Field of FMP Projected against radial vector kernel. These are extracted at Keypoints 12 and 13 of Figure (2). Vertical lines indicate local minima. The nested blobs give rise to two local minima, the single blob, to only one.

The locations of minima are sought in each profile, and this is used to identify the most appropriate scales for keypoint characterisation, illustrated by circles in Figure (2).

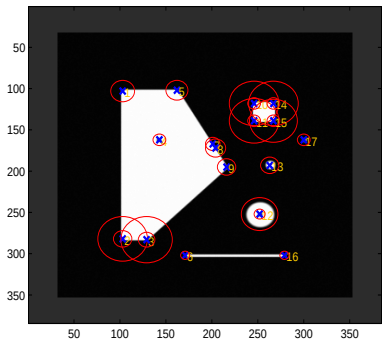

Fig. 2. FMP Keypoints and associated scales for phantom image.

3. AUGMENTING FMP KEYPOINTS: CM KEYPOINTS

CM (Circular Measure) Keypoints have been introduced to augment the FMP keypoints. The argument is as follows: FMP keypoints find points of relatively high curvature in the image. However, it is arguable that regions which are *surrounded* by well-defined boundaries might also be useful in identifying location correspondence. To achieve this, we generate the orientation dominance field described by [5] from the Q-Shift DTCWT complex filter outputs. This field is a simple simulation of neuronal encoding performed by complex

cells in primate visual cortex, and mimics divisive normalisation [13]. Our implementation is defined as

$$\mathbf{O}^{(\ell)}(m,n) = \frac{\sum_{k=1}^{6} |f_k^{(\ell)}| e^{2j\phi_k}}{\epsilon + (\sum_{k=1}^{6} |f_k^{(\ell)}|^2)^{1/2}} \tag{4}$$

where ϵ is a conditioning constant set to 0.04 of the peak image intensity. ϕ_k provides approximate directions of tuning of the band-pass wavelet channels. Because one cannot *a priori* identify what the scale, contrast, or pattern/shape of such boundaries might be useful keypoints, we merely seek to group circular arrangements of this orientation dominance field to yield CM keypoint responses at several scales. The accumulation of this orientation vector information from a region around each pixel is performed by using complex convolution with a vector kernel defined with magnitude as in Equation (2), but with the orientation field of (3) rotated by 90°. Only one radius is selected, with peak response at $5 \times 2^{\ell}$ pixels. This operation also yields response spaces, $C^{(\ell)}$ that may be considered as accumulator spaces[1]. Examples of these responses, at 5 different scales of the DTCWT decomposition of a white square, shown in Figure (3(a)), are presented in the remainder of Figure 3 below.

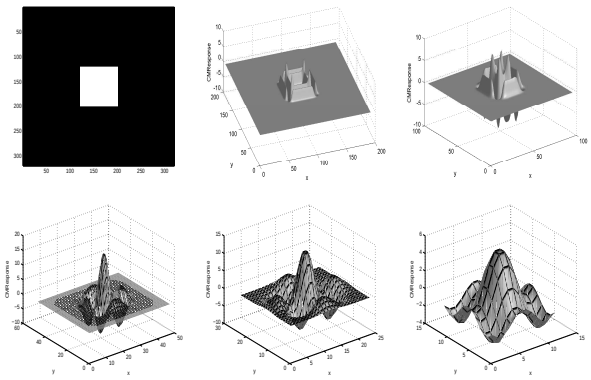

Fig. 3. The Circular Measure (CM) response of the image corresponding to a white square on a dark background at scales $\ell = 1, 2, .., 5$

From early experiments, it does not appear that the CM keypoints suffer from the problem of falling onto ridges (by virtue of the "action at a distance" form that the grouping kernel takes). Thus, they are not subjected to the ridge elimination procedure that the FMP keypoints are.

3.1. Intrinsic Scale for CM KeyPoints

Rather than define precise keypoint scales for the CM Keypoints, we assign an approximate scale, which is simply derived from the associated DTCWT at which a local maximum (peak) has been recorded. The phantom image introduced in [1], and with both types of keypoints and associated scales, is shown in Figure (4)

Note also that quantisation of the CM scales is performed rather coarsely, and room for improvement remains: it would be sensible, within each scale, to apply several radii of boundary response grouping kernels. This will be done in future work.

[1]The nature of the term accumulator is here inherited from the similarity that the grouping has to the Compact Hough Transform, and is not related to the FMP multi-scale accumulation used above.

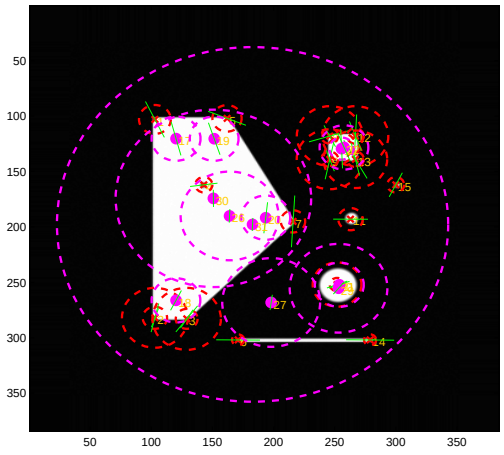

Fig. 4. FMP and CM Keypoint locations and scales; green lines indicate direction of wavelet energy at the single keypoint location in wavelet space. FMP keypoints are represented with × markers, CM keypoints are denoted by filled magenta circles.

4. PHASE INVARIANCE OF DETECTION

The FMP and CM Keypoints both contain invariance to phase and local image symmetry. In Figure (5), we compare the keypoints in a tilted, solid square under dramatic changes in image contrast and local phase. Both the FMP and CM keypoints are quite stable under contrast inversion, as shown in the top row of Figure (5). The CM keypoints also have additional stability to local gross phase changes, a feature that may be useful in shape primitive matching; Figures (5(c)) and (d) show the keypoints detected when the perimeter of the square is used instead. The location and scales of the CM keypoints change, but still locate, approximately, the corners and centre of the square. The FMP keypoints in this case have fallen below the firing threshold. The CM keypoint location strategy, although currently suffering from inaccuracies at coarse scales, thus does have the interesting property of robustness to quite gross phase changes. Note that the neither the solid nor outline square have been band-limited, and each were generated by rotating a square binary image using standard bilinear interpolation and perimeter finding tools.

Caution needs to be exercised in suggesting that the property of phase invariance always a good one to have in a keypoint *descriptor*, as opposed to a detector. If high specificity of local image phase is required, then phase invariance has the cost of loss of such discriminability. However, phase invariance *is* likely to be an important property for locating stable keypoints that lead to accurate shape comparisons.

5. CONCLUSIONS

The use of keypoints constructed from the DTCWT appears promising when drastic illumination changes might be possible. The phase invariance of the divisive neuronal model appears to lead to the desired effect of similar keypoints being created from shape boundaries as from solid shapes.

In particular, the keypoint locations appear to be intuitive, and the augmentation of the product of filter outputs proposed in [1] by

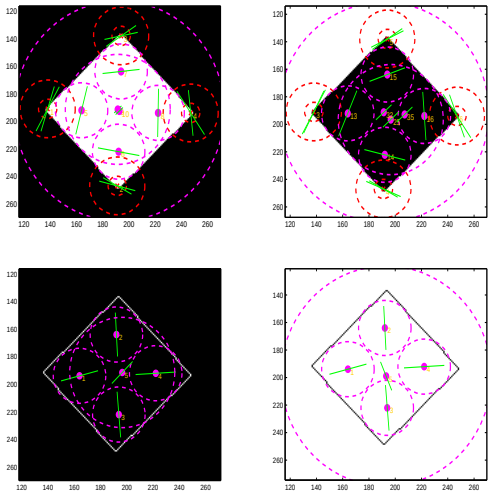

Fig. 5. Contrast and Phase stability of Keypoint detection. Note slight inaccuracies in the CM keypoints due to the localisation in decimated wavelet space.

the CM keypoints generates keypoints in the interiors of otherwise featureless shapes in synthetic images. Both classes of keypoints appear reasonable when applied to indoor scenes, and scenes containing people. A high proportion of keypoints appear either on facial features, or centered on faces. Thus, both classes of keypoints appear suitable for application as a focus-of-attention mechanism for identifying probable locations of facial detail in scenes. This may be used to guide algorithms to facial-like features for directing attention in adaptive vision systems.

This paper has not addressed the construction of a descriptor based on the DTCWT, although suggestions have been made by [4]. Further work is being carried out to attain sub-pixel keypoint localisation, to reduce computational effort for the CM keypoints, and to construct shape descriptors from the orientation dominance field of [5].

6. REFERENCES

[1] J. Fauqueur, N. Kingsbury, and R. Anderson, "Multiscale keypoint detection using the dual tree complex wavelet transform," in *IEEE International Conference on Image Processing*, 2006.

[2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. [Online]. Available: citeseer.ist.psu.edu/lowe04distinctive.html

[3] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, no. 3, pp. 234–253, May 2001.

[4] N. Kingsbury, "Rotation invariant local feature matching with complex wavelets," in *EUSIPCO*. Paper ID 1568982135, 4-8 September 2006.

[5] A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 14, pp. 948 – 959, July 2005.

[6] A. A. Bharath, "A tiling of phase-space through self-convolution," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3581–3585, December 2000.

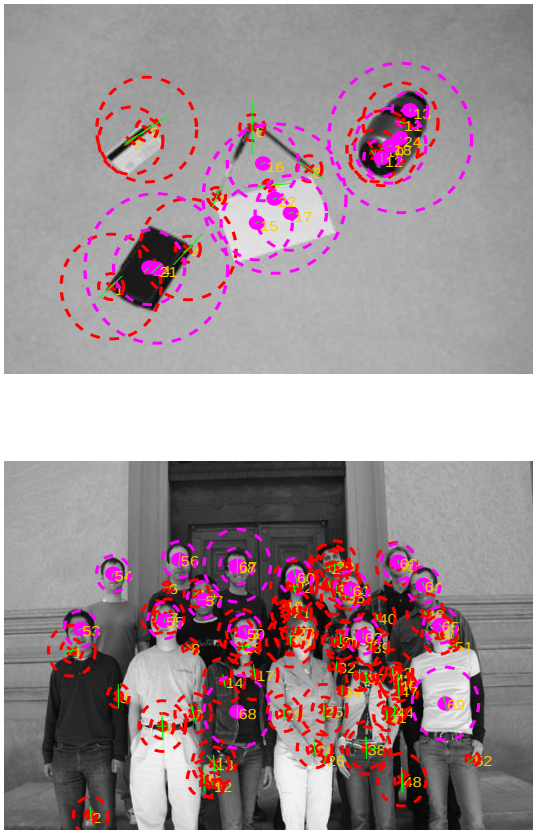

Fig. 6. Top, Keypoints obtained from indoor scene of objects on textured rug. (b) Bottom, Keypoints obtained for an image containing faces.

[7] T. Lindeberg, "Principles for automatic scale selection," in *Handbook on Computer Vision and Applications*. Academic Press, 1999, vol. 2, pp. 239–274.

[8] T. Kadir and M. Brady, "Scale, saliency and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, November 2001.

[9] T. Lindeberg and L. Bretzner, "Real-time scale selection in hybrid multi-scale representations," in *Scale Space 2003*, ser. Lecture Notes in Computer Science, vol. 2695, 2003.

[10] J. Crowley and O. Riff, "Fast computation of scale normalised Gaussian receptive fields," in *Scale Space 2003*, ser. Lecture Notes in Computer Science, vol. 2695, 2003, pp. 584–598.

[11] A. A. Bharath, "Scale steerability of a class of digital filters," *Electronics Letters*, vol. 34, no. 11, pp. 1087–1088, 1998.

[12] J. Ng and A. Bharath, "Steering in scale space to optimally detect image structures," *Lecture Notes In Computer Science (LNCS)*, vol. 3021, pp. 482–494, 2004.

[13] E. Simoncelli and D. Heeger, "A model of neuronal responses in visual area MT," *Vision Research*, vol. 38, no. 5, pp. 743–761, 1998.