Enhanced shift and scale tolerance for rotation invariant polar matching with dual-tree wavelets

J. D. B. Nelson and N. G. Kingsbury

***Abstract*—Polar matching is a recently developed shift and rotation invariant object detection method that is based on dual-tree complex wavelet transforms or equivalent multiscale directional filterbanks. It can be used to facilitate both keypoint matching, neighbourhood search detection, or detection and tracking with particle filters. The theory is extended here to incorporate an allowance for local spatial and dilation perturbations. With experiments, we demonstrate that the robustness of the polar matching method is strengthened at modest computational cost.**

I. Introduction

Some well known important wavelet properties include (bi)orthogonality, Besov regularity, compact support, and symmetry. Commonly however, object detection problems require the consideration of extra properties because two objects are often defined to be in the same class if one object is similar to some transformation of the other. If wavelets are to be used for object detection tasks, then either the objects must somehow be normalised first, or the wavelet coefficients must be invariant to certain transformations.

In practice, normalisation can be difficult. For translation invariance, some previous works have implemented a variant of the 'spin-cycle' method of Coifman and Donoho [1] whereby extra training samples are created by shifting the original ones. A more elegant method is to construct transforms which are themselves invariant.

The shiftable wavelet, introduced by Simoncelli et al [10], satisfies a slightly weaker condition than shift invariance but is less redundant than the spin cycle. The dual-tree complex wavelet transform (DTCWT), introduced by Kingsbury [4], [9], has good shift invariance and offers even lower redundancy with greater computational efficiency. Moreover, a recent extension of the DTCWT, known as polar matching [5], also possesses approximate rotation invariance. Unlike earlier DTCWT rotation invariant work of Hill et al [3], polar matching retains the phase information of the complex coefficients and therefore represents a richer descriptor.

Polar matching applications include keypoint matching, neighbourhood search detection, and detection and tracking with particle filters [7]. Since keypoints will not necessarily be centred on exactly the same object components from one image to the next, robustness to small displacement errors can be the key to the success of the method. In neighbourhood search detection, robustness to shifts allows the polar matching features to be sampled more sparsely. Added shift tolerance for particle filter polar matching would mean that the location of the particles would be less critical. This can add robustness to the observational model or allow fewer particles to be used. In all of these applications, scale tolerance would add robustness to unknown changes in distance between sensor and target.

J. D. B. Nelson and N. G. Kingsbury are with the University of Cambridge.

In this paper we shall consider the question of how to efficiently incorporate an allowance for an unknown local spatial and dilation perturbation into the polar matching method. The next section summarises the basic polar matching method. Section III introduces a shift tolerant extension to polar matching. This is further extended to include scale tolerance in Section IV. Experiments on a 73 image dataset from Caltech are described in Section V. Finally, conclusions are drawn in Section VI.

II. Polar matching

At each scale level, the DTCWT decomposes an image into six subbands. Each subband coefficient can be thought of as a response to a bandpass directional filter at a particular location. Together with their complex conjugates, the coefficients constitute 12 different directions, regularly spaced at $(30k-15)^\circ$, for $k = 1, \ldots, 12$. For the purposes of polar matching the 2-d real and imaginary impulse responses in the 45° and 135° directions are modified, as described in [5] and depicted in Figure 1, to have centre frequencies that match those of the other directions or subbands. In addition, the phases of the 6 band outputs are all centred to zero by a simple multiplication of $\{j, -j, j, -1, 1, -1\}$ respectively. In doing so, 6 opposing directions can be obtained by conjugating the 6 complex subband coefficients.

As illustrated in Figures 2 and 3, the elementary form of the polar matching method samples these 6 subband coefficients at 12 points around a circle and at one point at the circle centre. The coefficients are then assembled into what is known as a polar matching matrix (P-matrix), thus:

$$\mathbf{P} = \begin{bmatrix} m_1 & j_1 & k_1 & l_1 & a_1 & b_1 & c_1 \\ m_2 & i_2 & j_2 & k_2 & l_2 & a_2 & b_2 \\ m_3 & h_3 & i_3 & j_3 & k_3 & l_3 & a_3 \\ m_4 & g_4 & h_4 & i_4 & j_4 & k_4 & l_4 \\ m_5 & f_5 & g_5 & h_5 & i_5 & j_5 & k_5 \\ m_6 & e_6 & f_6 & g_6 & h_6 & i_6 & j_6 \\ m_1^* & d_1^* & e_1^* & f_1^* & g_1^* & h_1^* & i_1^* \\ m_2^* & c_2^* & d_2^* & e_2^* & f_2^* & g_2^* & h_2^* \\ m_3^* & b_3^* & c_3^* & d_3^* & e_3^* & f_3^* & g_3^* \\ m_4^* & a_4^* & b_4^* & c_4^* & d_4^* & e_4^* & f_4^* \\ m_5^* & l_5^* & a_5^* & b_5^* & c_5^* & d_5^* & e_5^* \\ m_6^* & k_6^* & l_6^* & a_6^* & b_6^* & c_6^* & d_6^* \end{bmatrix}$$

where the subscripts k determine the subband orientations $(30k-15)^\circ$, the coefficients labelled m are taken from the

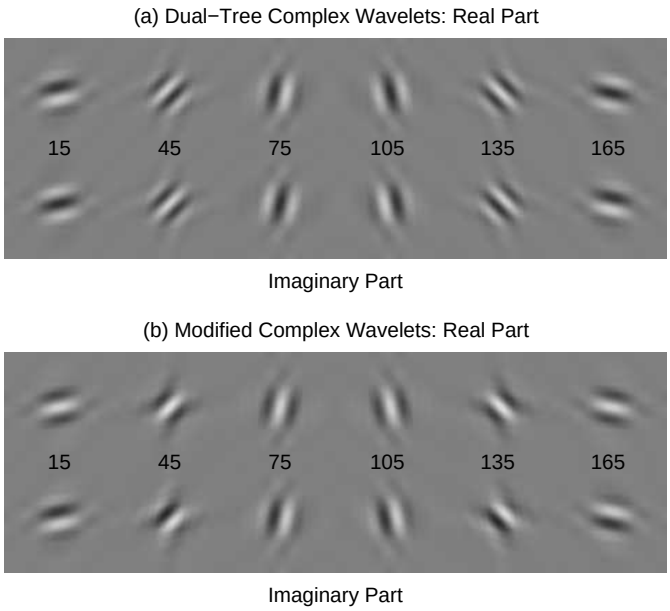


Fig. 1. In the interests of rotation invariance, the DTCWT 2-d real and imaginary impulse responses in the $45°$ and $135°$ directions are modified. Taken from [5].

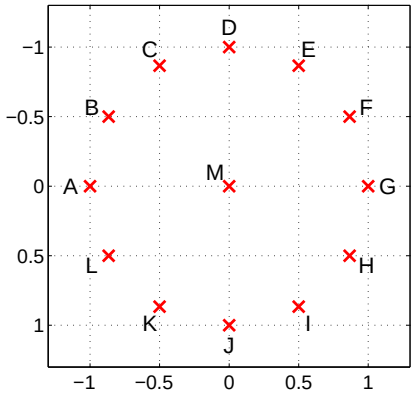


Fig. 2. Sampling locations of the DTCWT coefficients for the polar matching method. Taken from [5].

midpoint, and the coefficients $a, b, \ldots l$ determine the locations of the sample points as in Figure 2. The arrangement of the DTCWT coefficients ensures that each $30°$ rotation of the image about the centre point of the sampling circle produces a cyclical shift of each of the columns of the polar matching matrix (P-matrix).

Given two images, one a $30k°$ rotation of the other, a summation of the column-wise correlations between the two P-matrices will give a response curve, with respect to rotation angle, and a maximum at a shift of k. Hence the location of the correlation peak can be used to estimate the difference in orientation angle between two similar objects. Furthermore, the estimate can be improved by performing the correlation as a zero padded dot product in the Fourier domain before using an inverse FFT to obtain an upsampled correlation result. Typically, the original 12 samples are extended by a factor of 4 to obtain $7.5°$ rotational spacing.

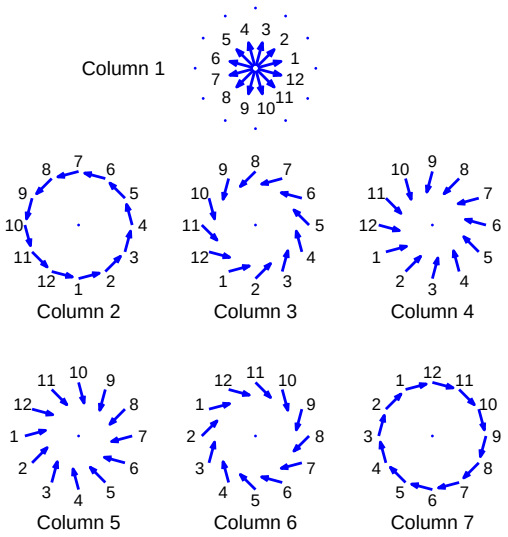


Fig. 3. Locations and orientations of the DTCWT coefficients and how they are arranged in the polar matching matrix (P-matrix). Each orientation describes a coefficient, or conjugate, of one of the six subbands. The numbers denote the orientation and the P-matrix row position. Taken from [5].

The part of the spectrum to be zero padded should be tailored differently to suit each column of the P-matrix. In particular, consider the P-matrix formed at the centre of a rotation of a single step edge. As the edge is rotated, the response of column 1 will vary as a lowpass function. Columns 2 and 7 will vary slightly quicker as bandpass functions, 3 and 6 quicker still, and 4 and 5 the quickest as highpass functions. The rate of change depends on the subband orientation with respect to the radial direction. Denoting this angle by α, and referring to Figure 3, we have $\alpha = 75°$ for columns 2 and 7, $45°$ for 3 and 6, and $15°$ for 4 and 5. Generally, the centre frequency of the columns is proportional to $\cos\alpha$.

Now consider the DTCWT decomposition of a step edge orientated at $15°$ to the horizontal, placed in the centre of a 128×128 subimage. Figure 4 shows that the phase response of the centre coefficient at the 4th finest scale level taken from subband 1 (oriented such that the stripe direction is parallel to the edge direction) shifts by almost π radians over a displacement of 12 pixels (note that 12 pixels is equivalent to $12/2^4 = 3/4$ samples at the 4th finest scale level). Hence, over one sample at the 4th finest scale level, the phase will shift by $4\pi/3$ radians. Therefore, the rate of phase change with respect to a rotating step edge will be approximately equal to $\frac{4\pi}{3}\cos\alpha$.

After a dot product of the two P-matrices in the Fourier domain is carried out, the result is then subjected to zero padding, followed by an inverse Fourier transform. One then arrives at a response curve as a function of orientation, sampled at $7.5k°$, for $k = 0, \ldots, 47$. The P-matrix represents an approximately rotation invariant feature under the operation of correlation. The polar matching process transforms rotations in the original object into shifts in the feature space.

To enrich the descriptor, further scale levels, sampling circles, and colour channels may be considered by appending coefficients as extra columns of the P-matrix. For the experiments

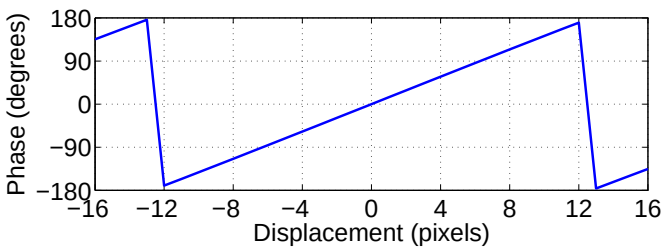


Fig. 4. A step edge orientated at 15° to the horizontal, and placed in the centre of a 128×128 subimage is decomposed by the DTWCT. The phase angle of the centre DTCWT coefficient from subband 1 at the 4th finest scale level is plotted with respect to vertical displacement. Note how the phase completes approximately 3/4 of a cycle over the sampling period of $2^4 = 16$ pixels.

carried out in this paper, two sampling rings and centre point from the third finest level are used, together with one ring and centre point at the fourth finest, and just the centre point at the fifth finest level. For simplicity, colour information was converted into monochrome values prior to any processing. This results in 13 columns of 12 coefficients from the third finest level, 7 columns from the fourth, and one column from the fifth. Hence the number of columns $L = 21$. Figure 5 illustrates the sampling locations used.

Consider the column-wise Fourier transform of a template P-matrix taken about some centre point. As stated above, in practice the 12 elements in each column are extended by a factor of 4 to generate $K = 48$ rows. Denote the (k, ℓ)th element of the resulting matrix by $h_{k,\ell}$. Likewise, let $f_{k,\ell}(\mathbf{x})$ be the elements in the column-wise Fourier transform of a test image P-matrix taken about the point $\mathbf{x}$. The polar matching operation between the two can be expressed as

$$g(\mathbf{x};\theta) = \Re\left\{\frac{1}{K}\sum_{k=0}^{K-1}\exp\left(\frac{2\pi i\theta k}{K}\right)\sum_{\ell=0}^{L-1}\chi_\ell\left[k\right]\overline{h}_{k,\ell}f_{k,\ell}(\mathbf{x})\right\}.$$

The element-by-element products between the columns of the two matrices are carried out in the second summation. Before the products are taken, the elements are multiplied by the indicator function $\chi\colon \mathbb{Z} \mapsto \{0,1\}$ which, for each column ℓ, takes zeros over an appropriate part of the Fourier domain and ones elsewhere; it can be seen as an ideal bandpass filter. The inverse Fourier transform is then carried out in the outer summation over k.

Since the column 1 coefficients vary as a lowpass function,

$$\chi_1\left[k\right] = \begin{cases} 1 & \text{if } 0 \le k \le 5, \text{ and } 42 \le k \le 47 \\ 0 & \text{if } 6 \le k \le 41 \end{cases}$$

Likewise, the centre points from any other scale level will be multiplied by χ_1. For columns 2 to 7, since the rate of phase change is approximately $\frac{4\pi}{3}\cos\alpha$, it follows that, for $K = 48$:

$$\begin{aligned} \chi_2\left[k\right] = \chi_7\left[k\right] &= \chi_1\left[(k+1) \bmod 48\right] \\ \chi_3\left[k\right] = \chi_6\left[k\right] &= \chi_1\left[(k+3) \bmod 48\right] \\ \chi_4\left[k\right] = \chi_5\left[k\right] &= \chi_1\left[(k+4) \bmod 48\right]. \end{aligned}$$

The inner rings at other levels will similarly be multiplied by $\chi_2, \ldots, \chi_7$. Since the outer ring at the third finest level (see Figure 5) has twice the radius of the inner ring, the rate of phase change will double. If the outer rings are assembled into columns 16 to 21, then

$$\begin{aligned} \chi_{16}\left[k\right] = \chi_{21}\left[k\right] &= \chi_1\left[(k+2) \bmod 48\right] \\ \chi_{17}\left[k\right] = \chi_{20}\left[k\right] &= \chi_1\left[(k+6) \bmod 48\right] \\ \chi_{18}\left[k\right] = \chi_{19}\left[k\right] &= \chi_1\left[(k+6) \bmod 48\right]. \end{aligned}$$

Generally,

$$\chi_\ell\left[k\right] = \chi_1\left[\left(k + \text{round}\left(\max\left\{\frac{4\rho\pi}{3}\cos\alpha_\ell, 6\right\}\right)\right) \bmod 48\right]$$

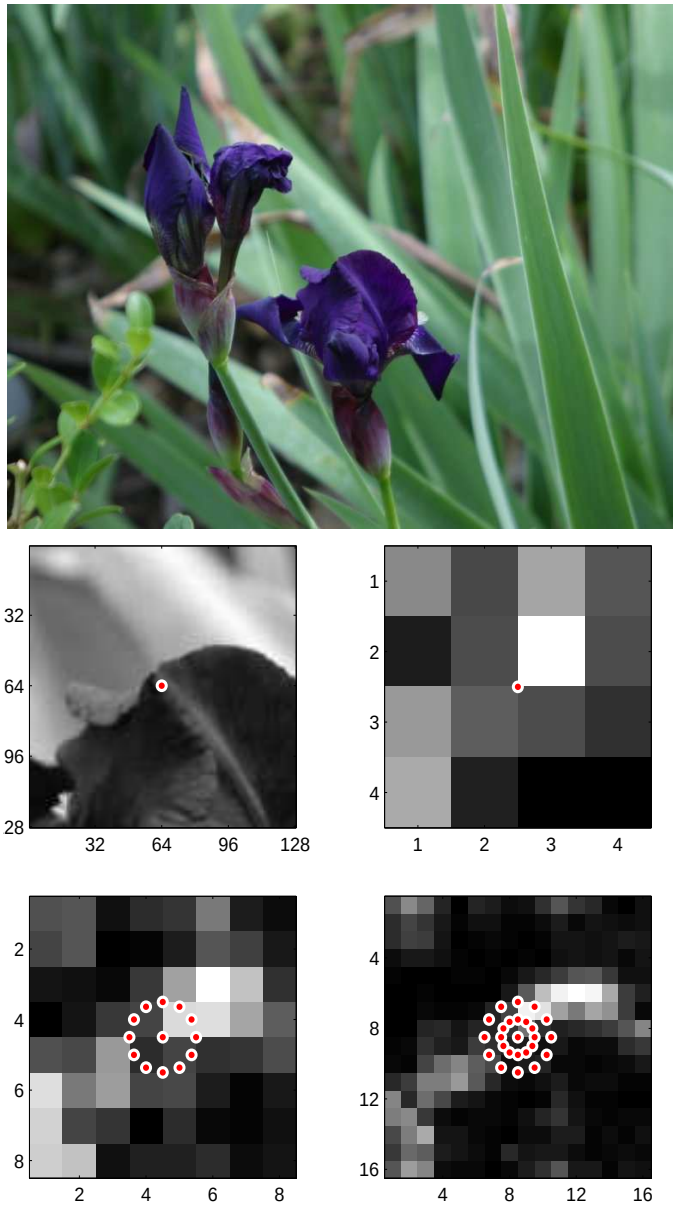


Fig. 5. Multiple scale level sampling locations of the DTCWT coefficients for the polar matching method. A 128×128 subimage is extracted from the centre of the image (top most) and decomposed into 3 levels of detail. Top left: original subimage; top right: 5th finest level; bottom left: 4th finest level; bottom right: 3rd finest level. For illustrative purposes, only the absolute values of the DTCWT coefficients in subband 1 are shown. In practice, the real and imaginary parts are used from all subbands. The original image is '201.jpg' from Caltech's 'PP_Toys_03' full resolution dataset [8].

where α_ℓ is the subband orientation with respect to the radial direction of the filters. Note that the maximum allowable shift of the function χ_1 is $k+6$ (any further shifts would make the high pass functions χ_{18} and χ_{19} become low pass functions). The radius of the sampling circle ρ is measured in samples of the respective coefficient space. In our experiments, the fourth finest level has a ring of unit radius, and the third finest level has an inner ring of radius 1 and outer ring of radius 2.

The fast Fourier transform can be used to speed up the polar matching computation. This only needs to be done once, rather than on every column. Alternatively, if there exists prior information about the orientation of the target, the Fourier transform need only be computed for a subset of θ. For example, when tracking objects using video it is reasonable to assume that the target orientation rate of change is bounded. Finally, the real component $\Re$ is taken in order to return the purely real correlation intensity. For simplicity, and to aid the development of the shift and scale tolerant polar matcher in the next section, this is first rewritten as

$$g(\mathbf{x};\theta) = \Re\left\{\sum_{k=0}^{K-1}\sum_{\ell=0}^{L-1} w_{k,\ell}(\theta)\overline{h}_{k,\ell} f_{k,\ell}(\mathbf{x})\right\},$$

where

$$w_{k,\ell}(\theta) = \frac{1}{K}\exp\left(\frac{2\pi i\theta k}{K}\right)\chi_\ell\,[k]\ .$$

Now put $j = k + K\ell$, and $n = KL$. Define the column vector $\mathbf{h} = (h_j)_0^{n-1} \in \mathbb{C}^n$ as

$$\mathbf{h} \triangleq [h_{0,0},\ldots,h_{K-1,0},h_{0,1},\ldots,h_{K-1,1},\ldots h_{K-1,L-1}]^T\ .$$

That is, $\mathbf{h}$ concatenates the columns of the matrix $(h_{k,\ell})$. Likewise, form $\mathbf{w}$ and $\mathbf{f}$. We then have

$$g(\mathbf{x};\theta) = \Re\left\{\sum_{j=0}^{n-1} w_j(\theta)\overline{h}_j f_j\right\},$$

which is now just a weighted inner product, namely $\Re\left\{\mathbf{h}^H\mathbf{W}_\theta\mathbf{f}(\mathbf{x})\right\}$, where the superscript ^{H} denotes complex conjugate transpose. For each θ the matrix $\mathbf{W}_\theta = \mathrm{Diag}(\mathbf{w}(\theta))$ is diagonal. This summation only takes place over $n/4 = 252$ non-zeros because of the Fourier domain zero padding.

To summarise, the polar matching operation between the template Fourier P-vector $\mathbf{h}$ and test image Fourier P-vector $\mathbf{f}$ about the point $\mathbf{x}$ is

$$g(\mathbf{x};\theta) \triangleq \big(\mathbf{f}(\mathbf{x}) \star \mathbf{h}\big)(\theta) \triangleq \Re\left\{\mathbf{h}^H\mathbf{W}_\theta\mathbf{f}(\mathbf{x})\right\}\ .$$

Since both $\mathbf{h}$ and $\mathbf{W}_\theta$ are independent of the test image, the product $\mathbf{h}^H\mathbf{W}_\theta$ can be precomputed and stored.

In the following, the polar matching approach is strengthened so that a P-matrix constructed from a centre point location $(x+\delta x, y+\delta y)$ and scale $s+\delta s$ will still obtain a large correlation score when matched with the same image centred on (x,y) at scale s.

Several benefits of shift and scale tolerant polar matching can be readily realised. One such application is keypoint matching.

In a similar way to SIFT [6], DTCWT keypoints can be established in scale and space [2]. Contrary to SIFT, the polar keypoint matching method does not choose a dominant orientation for each keypoint but rather makes use of correlation scores at all rotations. Since keypoints will not necessarily be centred on exactly the same object components from one image to the next, robustness to small displacement errors can be the key to the success of the method.

A second application is to use polar matching in a template matching approach for object detection in video [7]. Assuming that we have access to one or more examples of the target or object of interest, we can use polar matching to search an unknown test image by extracting polar matrices from some neighbourhood or window in the test image and correlating each one with the template stored in a database. Hence, a correlation surface can be obtained. For each θ, the steepness of the correlation surface about the maximum in $\mathbb{R}^2$ can be controlled to some extent by the template size and the choice of wavelet decomposition levels used. However, the size of the template relative to image size will be determined by the application, data, and object of interest. In practice, the correlation surface is computed over a discretised set of points. A full, exhaustive computation would involve forming test P-matrices at every pixel, or perhaps sub-pixel, in the search region. If such an approach proves intractable then it becomes necessary to calculate the correlation surface over a sparser set of locations. On the other hand, a sparser search carries with it the risk of missing the correlation peak altogether. In this setting, robustness to shifts allows the P-matrices to be sampled more sparsely. Scale tolerance adds robustness to changes in distance between sensor and target. In the same context, particle filtering has also been used with polar matching [7]. Here, polar matching scores are computed at each particle location to form the observational model. With added shift tolerance the location of the particles becomes less critical and can either add robustness to the observational model or allow fewer particles to be used.

III. Shift tolerance

In this section we present an extension to the original polar matching method to incorporate an allowance for larger local spatial displacement errors. In doing so, it will be seen that the steepness of the correlation surface about the maximum is mitigated. As above, let $\mathbf{h}$ and $\mathbf{f}$ be the Fourier transforms of the polar matching features of the template sub-image and test image, respectively. As such, they can be considered mappings from the original pixel coordinates to the n dimensional complex hypersphere $S^n = \{\mathbf{x} \in \mathbb{C}^n\colon\ \|\mathbf{x}\| = 1\}$. That is

$$\mathbf{h},\mathbf{f}\colon \mathbb{R}^2 \mapsto S^n.$$

Polar matching, denoted by $\star$, is the symmetric bilinear operator

$$\star\colon S^n \times S^n \mapsto [-1,1]^m \subset \mathbb{R}^m,$$

from the complex n-sphere to the real m-cube, where m is the number of bins for θ. Recall that it is defined as the weighted inner product:

$$(\mathbf{f}\star\mathbf{h})(\theta) \triangleq \Re\left\{\mathbf{h}^H\mathbf{W}_\theta\mathbf{f}\right\} \triangleq \Re\left\{\mathbf{f}^H\mathbf{W}_\theta\mathbf{h}\right\}\ .$$

Now define $\Delta\mathbf{h} = \mathbf{h}(\cdot+\Delta\mathbf{x}) - \mathbf{h}$. Then, for small $\Delta\mathbf{x}$, assume that a small change in $\mathbf{h}$ is linear with respect to small spatial

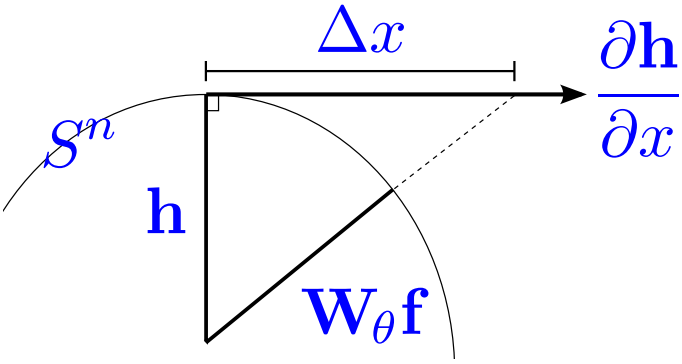


Fig. 6. The shift tolerance problem, simplified in 2-D space. Find the shift distance $\Delta\mathbf{x}$ along the surface of the hypersphere S^n, such that the unit vector $\mathbf{h}$ is rotated into $\mathbf{W}_\theta\mathbf{f}$.

shifts. That is

$$\Delta\mathbf{h} = \mathbf{J}\Delta\mathbf{x}, \qquad \mathbf{J} = \left[\frac{\partial\mathbf{h}}{\partial x}, \frac{\partial\mathbf{h}}{\partial y}\right] \in \mathbb{C}^{n\times 2}.$$

This is equivalent to a first order Taylor series expansion of $\mathbf{h}(\cdot + \Delta\mathbf{x})$. Then

$$\begin{aligned} \mathbf{f} \star \mathbf{h}(\cdot + \Delta\mathbf{x}) &= \mathbf{f} \star (\mathbf{h} + \Delta\mathbf{h}) \\ &= \mathbf{f} \star (\mathbf{h} + \mathbf{J}\Delta\mathbf{x}) \\ &= \Re\, \mathbf{f}^H \mathbf{W}_\theta \left(\mathbf{h} + \mathbf{J}\Delta\mathbf{x}\right). \end{aligned}$$

As Figure 6 illustrates, the problem that shift tolerant polar matching addresses is to find the maximum shifted polar matching correlation score with respect to an unknown small shift $\Delta\mathbf{x}$. Since $\mathbf{h}$ and $\mathbf{f}$ are both normalised to unit length, the aim is to get as close to $\mathbf{W}_\theta\mathbf{f}$ by moving a distance of $\Delta\mathbf{x}$, from $\mathbf{h}$, along the surface of the hypersphere in a direction orthogonal to the radial vector $\mathbf{h}$. That is

$$\mathbf{x}_\theta = \underset{\Delta\mathbf{x}}{\operatorname{argmax}}\, \Re\left\{\mathbf{f}^H \mathbf{W}_\theta \frac{\mathbf{h} + \mathbf{J}\Delta\mathbf{x}}{\|\mathbf{h} + \mathbf{J}\Delta\mathbf{x}\|}\right\}, \quad \mathbf{J}^H\mathbf{h} = \mathbf{0}. \quad (1)$$

Ideally, we want $\mathbf{x}_\theta$ such that

$$\mathbf{h} + \mathbf{J}\mathbf{x}_\theta = \mathbf{W}_\theta\mathbf{f}, \quad \mathbf{J}^H\mathbf{h} = \mathbf{0}. \quad (2)$$

However, because S^n is an n) dimensional space and $\mathbf{x}_\theta$ has only 2 dimensions, this is an overdetermined set of equations. Instead, the solution, in the least squares sense, is

$$\begin{aligned} \mathbf{x}_\theta &= \Re\left\{\left(\mathbf{J}^H\mathbf{J}\right)^{-1}\right\} \Re\left\{\mathbf{J}^H\mathbf{W}_\theta\mathbf{f}\right\} \\ &= \Re\left\{\left(\mathbf{J}^H\mathbf{J}\right)^{-1}\right\} (\mathbf{J}\star\mathbf{f})^T(\theta). \end{aligned} \quad (3)$$

Note that, since $\|\mathbf{h}\| = 1$ and $\mathbf{h}^H\mathbf{J} = 0$, the denominator of (1) is

$$\|\mathbf{h} + \mathbf{J}\Delta\mathbf{x}\| = \sqrt{1 + \mathbf{x}_\theta^H \Re\{\mathbf{J}^H\mathbf{J}\}\mathbf{x}_\theta}\,.$$

Hence, substituting $\mathbf{x}_\theta$ from (3) into the right hand side of (1) gives shift tolerant polar matching:

$$g^+(\mathbf{x};\theta) \triangleq \frac{(\mathbf{h}\star\mathbf{f}) + (\mathbf{J}\star\mathbf{f})\,\mathbf{A}\,(\mathbf{J}\star\mathbf{f})^T}{\sqrt{1 + (\mathbf{J}\star\mathbf{f})\,\mathbf{A}\,(\mathbf{J}\star\mathbf{f})^T}}, \quad (4)$$

where $\mathbf{A} = \Re\left\{\left(\mathbf{J}^H\mathbf{J}\right)^{-1}\right\}$. The term $\mathbf{J}^H\mathbf{J}$ is a 2×2 matrix and is independent of the test image. Under the reasonable assumption that $\partial\mathbf{h}/\partial x$ and $\partial\mathbf{h}/\partial y$ are linearly independent, $\mathbf{J}^H\mathbf{J}$ is a positive definite matrix and is therefore invertible. All other terms involve polar matching operations on the test image with the template and two spatial derivatives of the template. Therefore, the template $\mathbf{h}$, Jacobian $\mathbf{J}$, and matrix $\mathbf{A}$ should be pre-computed and stored in memory.

Compared with the original polar matching method, which just requires computation of $\mathbf{h}\star\mathbf{f}$, this shift tolerant version also requires $\mathbf{J}\star\mathbf{f} = [\partial\mathbf{h}/\partial x, \partial\mathbf{h}/\partial y]\star\mathbf{f}$. This comprises two weighted inner products, each of similar complexity to $\mathbf{h}\star\mathbf{f}$. Alternatively, using the inverse fast Fourier transform to compute the polar matching operations, we now require three FFTs (rather than one) per test location $\mathbf{x}$. Since an upsampling of 4 is used, these are 48-point FFTs. Once $\mathbf{J}\star\mathbf{f}$ has been computed, there is a minor additional cost: for each $\theta = 0, \ldots, 47$ and $\mathbf{x}$, of 6 multiplications and 2 adds to compute the quadratic form $(\mathbf{J}\star\mathbf{f})\,\mathbf{A}\,(\mathbf{J}\star\mathbf{f})^T$, one add for the numerator, one add and a square root for the denominator, and a division.

However, the overall computation of both the original and tolerant method is dominated by calculating the features $\mathbf{f}$ at each point $\mathbf{x}$. This overhead involves a DTCWT to decompose the test image. More crucially, at each $\mathbf{x}$ it also requires several bandpass interpolations to generate the coefficients around the sampling circles. As we will illustrate with some experiments in Section V, this is a worthwhile cost in order to reduce the sensitivity to displacement error.

IV. Scale tolerance

Shift tolerant matching can be extended quite naturally to shift and scale tolerant matching. We introduce a dilation variable ψ, such that

$$\mathbf{h}(\mathbf{x};\psi) \triangleq \mathbf{h}(\psi\mathbf{x}).$$

Now

$$\mathbf{h}\colon \mathbb{R}^3 \mapsto \mathbb{C}^n.$$

Define $\Delta\mathbf{h} \triangleq \mathbf{h}(\mathbf{x} + \Delta\mathbf{x}; \psi + \delta\psi) - \mathbf{h}(\mathbf{x};\psi)$. Then, for small $\Delta\mathbf{x}$ and $\delta\psi$, assume

$$\Delta\mathbf{h} = \mathbf{J}\begin{bmatrix}\Delta\mathbf{x}\\ \delta\psi\end{bmatrix}, \quad \mathbf{J} = \left[\frac{\partial\mathbf{h}}{\partial x}, \frac{\partial\mathbf{h}}{\partial y}, \frac{\partial\mathbf{h}}{\partial\psi}\right] \in \mathbb{C}^{n\times 3}.$$

I.e. this is a first order Taylor series expansion of the shifted and dilated template $\mathbf{h}(\mathbf{x} + \Delta\mathbf{x}; \psi + \delta\psi)$. Similar to the shift tolerant case, we want

$$\mathbf{h} + \mathbf{J}\begin{bmatrix}\mathbf{x}_\theta\\ \psi_\theta\end{bmatrix} = \mathbf{W}_\theta\mathbf{f}, \quad \mathbf{J}^H\mathbf{h} = \mathbf{0}.$$

The solution, in the least squares sense, is

$$\begin{aligned} \begin{bmatrix}\mathbf{x}_\theta\\ \psi_\theta\end{bmatrix} &= \Re\left\{\left(\mathbf{J}^H\mathbf{J}\right)^{-1}\right\} \Re\left\{\mathbf{J}^H\mathbf{W}_\theta\mathbf{f}\right\} \\ &= \Re\left\{\left(\mathbf{J}^H\mathbf{J}\right)^{-1}\right\} (\mathbf{J}\star\mathbf{f})^T(\theta). \end{aligned}$$

The shift and scale tolerant polar matching takes a similar form to (4). The only difference is that $\partial\mathbf{h}/\partial\psi$ has been appended as an extra column to $\mathbf{J}$.

TABLE I
ENERGY IMPROVEMENT RATIOS

method/experiment	mean	std
$\mathcal{M}g^{+}_{0,1}(\mathbf{x};0)$	3.02	0.53
$\mathcal{M}g^{+}_{\theta,1}(\mathbf{x};\theta)$	2.39	0.53
$\mathcal{M}g^{++}_{0,1}(\mathbf{x};0)$	4.62	1.91
$\mathcal{M}g^{++}_{0,\psi}(\mathbf{x};0)$	2.30	0.77

V. EXPERIMENTS

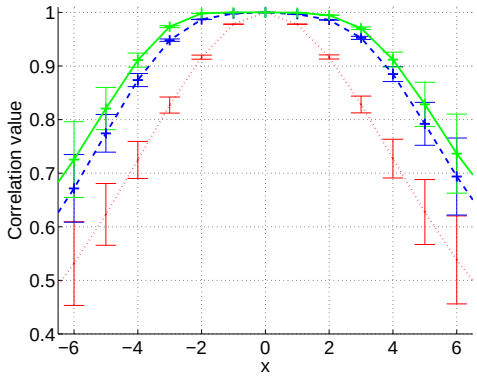
Caltech's 'PP_Toys_03' full resolution dataset [8] was used to investigate the effectiveness of the shift and scale tolerant methods. The 73 image dataset mostly comprises various toys against a grassy or stony background. To simplify reproducibility of results, no attempt was made to find suitable template centre points. Instead, templates of size 128×128 pixels were simply taken from the centre of each image. For simplicity, prior to any further processing, the RGB values were converted to intensity via $0.3R + 0.6G + 0.1B$. Figure 5 illustrates the 3 levels of DTCWT sampling used.

Correlation surfaces were obtained by performing polar matching between the template and test image regions centred about the template. I.e. we compute $g(\mathbf{x};\theta) = \mathbf{h}(\mathbf{x}) \star \mathbf{h}(\mathbf{0})$ about a local neighbourhood of $\mathbf{x} = \mathbf{0}$. As discussed earlier, the polar matching output $g(x, y; \theta)$ is a function of space (x, y) and orientation θ. Denote the shift tolerant matcher output as g^{+}, and the shift and scale tolerant output as g^{++}. For each template, the test images are shifted, rotated, and dilated versions of the template. Rotations were performed on the test images to show that the rotational invariance property of polar matching is not diminished by the tolerant methods. For each rotation $\Theta = 0, 7.5°, \ldots, 90°$ and dilation $\Psi = 1, 1.05, \ldots, 1.5$, and for each of the 73 test images, we obtain a correlation output $g_{\Theta,\Psi}(x, y; \theta)$. The shift tolerant matcher was applied to the shifted and rotated test images to give the correlation surfaces $g^{+}_{\Theta,1}(x, y; \theta)$. The shift and scale tolerant matcher was applied to the shifted and dilated test images to give $g^{++}_{0,\Psi}(x, y; \theta)$. A sum over all values of the correlation surfaces that are above 90% of the theoretical maximum of 1 is computed for the original and tolerant matchers. A ratio of the tolerant matcher score over the original score is then computed for comparison. That is, for the shift tolerant matcher, the ratio

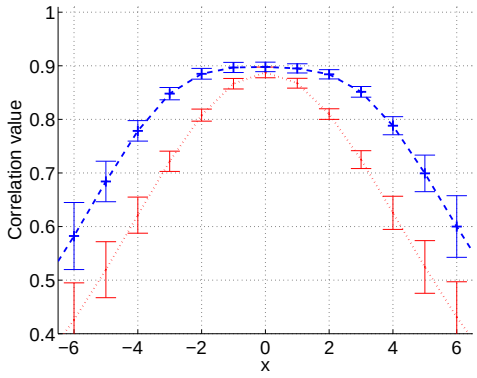
$$\mathcal{M}g^{+}_{\theta,1}(\mathbf{x};\theta) \triangleq \frac{\sum_{g^{+}>0.9} g^{+}_{\theta,1}(\mathbf{x};\theta)}{\sum_{g>0.9} g_{\theta,1}(\mathbf{x};\theta)},$$

gives a measure of the energy of the region within 0.9 of the height of the maximum. The mean and standard deviation of $\mathcal{M}$, taken over all experiments are given in Table I.

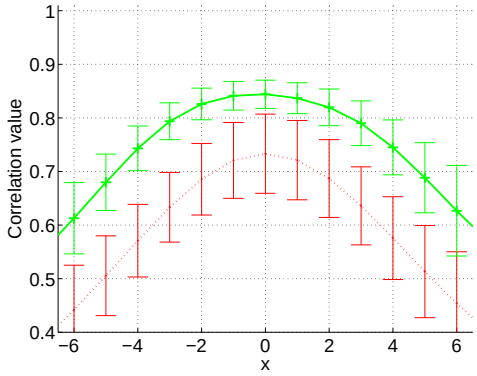
The behaviour with respect to non-targets was also investigated. Ten random points were taken from each of the 73 images at a minimum distance of 64 pixels from the centre. Each of these points was correlated with each of the 73 templates. The histograms of the resulting 53290 correlation scores are shown in Figure 8. By inspection, a small proportion of the image pairs in this experiment resemble scaled, rotated, and shifted versions of each other. Hence, the histogram overestimates the number of false positives. As might be



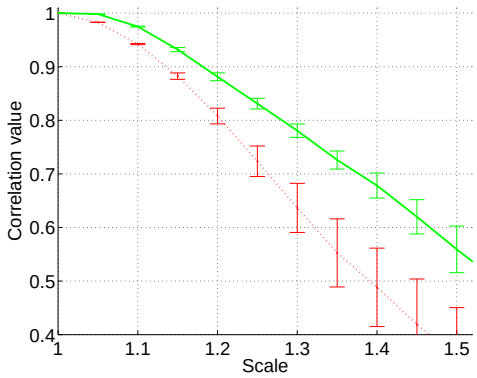
(a) $g_{0,1}(x, 0; 0)$



(b) $\underset{0\le\theta\le\pi}{\text{mean}}\, g_{\theta,1}(x, 0; \theta)$



(c) $\underset{\psi=1,1.05,\ldots,1.4}{\text{mean}}\, g_{0,\psi}(x, 0; 0)$



(d) $g_{0,\psi}(0, 0; 0)$

Fig. 7. Mean correlation values. With respect to x: (a) unrotated and undilated; (b) mean over rotations; and (c) mean over dilations. (d) Mean correlation values with respect to dilation. The original polar matching is plotted with a dotted red line, the shift tolerant method with a blue dashed line with crosses, and the shift and scale tolerant method with a crossed green solid line. The error bars denote variance over all experiments.

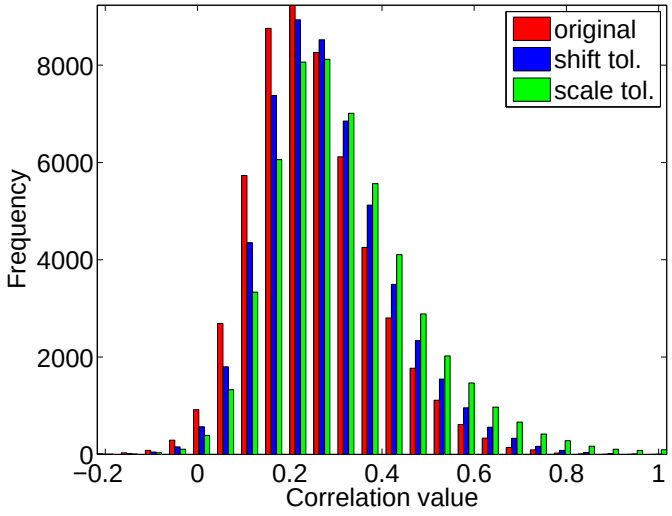


Fig. 8. Histograms of non-target correlation scores.

expected, the discriminative ability tends to diminish as more tolerance is added.

VI. Conclusion

Figure 7 and Table I show that the shift and scale tolerant methods behave as intended. Interestingly, Subplot (a) of Figure 7 shows that the shift-scale invariant methods are a little more tolerant to shifts than the shift invariant methods. This may be due to the fact that, for some image templates, a small perturbation in scale $(1+\delta)\mathbf{x}$ approximates a spatial shift. The results show that the test image needs to be scaled by a factor of more than 1.35 before the shift and scale tolerant correlation score falls below 70% of the maximum. If, however, the distance between sensor and known object is available then scale tolerance is less important. In this case, shift-rotational invariance and shift tolerance is still important and the shift tolerant polar matcher may be more appropriate.

In practice the Jacobian $\mathbf{J}$ is approximated by a finite difference. For example, the partial derivative with respect to x is estimated by

$$\frac{\partial \mathbf{h}}{\partial x} \approx \frac{1}{\epsilon}\big(\mathbf{h}(x+\epsilon, y) - \mathbf{h}(x, y)\big) \ .$$

The value $\epsilon = 1/10$ of a pixel was used in all experiments performed in this paper. It is conceivable that the finite difference increment ϵ could be optimised. Furthermore, the finite difference approximation could potentially be replaced with a more sophisticated discrete derivative.

It is also important to note that the shift and scale tolerant approach implied by (1) could be applied to any other correlation operation that can be expressed as a weighted inner product of the form $\mathbf{f}^H\mathbf{W}\mathbf{h}$. A simple example would be the classic matched filter with $\mathbf{W} = \mathbf{I}$, where $\mathbf{h}$ and $\mathbf{f}$ are simply intensity values of two images.

Investigations of shift and shift-scale tolerant matching to specific applications like keypoint matching and target detection and particle filter tracking in video should make for interesting further work.

References

[1] R. R. Coifman and D. Donoho. Wavelets and Statistics, Lecture notes in Statistics. In: *Translation-invariant de-noising. Antoniadis, A. and Oppenheim, G. (eds.)*, pages 125–150. Springer, 1995.

[2] J. Fauqueur, N. Kingsbury, and R. Anderson. Multiscale keypoint detection using the dual-tree complex wavelet transform. *Proc. IEEE Conference on Image Processing*, pages 1625–1628, 2006.

[3] P. R. Hill, D. R. Bull, and C. N. Canagarajah. Rotationally invariant texture features using the dual-tree complex wavelet tranform. In *Proc. ICIP*, 2000.

[4] N. G. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Journal of Applied and Computational Harmonic Analysis*, 10(3):234–253, 2001.

[5] N. G. Kingsbury. Rotation-invariant local feature matching with complex wavelets. In *Proc. European Conference on Signal Processing (EUSIPCO)*, pages 901–904, 2006.

[6] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[7] J. D. B. Nelson, S. K. Pang, S. J. Godsill, and N. G. Kingsbury. Tracking ground based targets in aerial video with dual-tree complex wavelet polar matching and particle filtering. *Fusion*, 2008.

[8] P. Perona. Caltech's 'PP_Toys_03' full resolution dataset. Website, 2003. http://www.vision.caltech.edu/pmoreels/Datasets/PP_Toys_03/FullResolution/TestSingleObjects.tar.

[9] 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.

[10] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Trans. Inform. Theory*, 38:587–607, 1992.