Phase-based Local Features

Gustavo Carneiro and Allan D. Jepson

Department of Computer Science
University of Toronto
{carneiro,jepson}@cs.toronto.edu

Abstract. We introduce a new type of local feature based on the phase and amplitude responses of complex-valued steerable filters. The design of this local feature is motivated by a desire to obtain feature vectors which are semi-invariant under common image deformations, yet distinctive enough to provide useful identity information. A recent proposal for such local features involves combining differential invariants to particular image deformations, such as rotation. Our approach differs in that we consider a wider class of image deformations, including the addition of noise, along with both global and local brightness variations. We use steerable filters to make the feature robust to rotation. And we exploit the fact that phase data is often locally stable with respect to scale changes, noise, and common brightness changes. We provide empirical results comparing our local feature with one based on differential invariants. The results show that our phase-based local feature leads to better performance when dealing with common illumination changes and 2-D rotation, while giving comparable effects in terms of scale changes.

Keywords: Image features, Object recognition, Vision systems engineering and evaluation, Invariant local features, Local phase information.

1 Introduction

View-based object recognition has recently received a great deal of attention in the vision literature. In this paper we are particularly interested in approaches based on local features (e.g. differential invariants in [20], and local scale-invariant features in [13]). These approaches have demonstrated their unique robustness to clutter and partial occlusion, while keeping the flexibility and ease of training provided by classical view-based approaches (see [15, 22]). However, to be successful for object recognition, local features must have the two properties: 1) be robust to typical image deformations; and 2) be highly distinctive to afford identity information.

We propose a novel local feature vector that is based on the phase and amplitude responses of complex-valued steerable filters. This builds on previous work [3] in which it was shown that the phase information provided by such filters is often locally stable with respect to scale changes, noise, and common brightness changes. Here we show it is also possible to achieve stability under rotation by selecting steerable filters.

The results of an empirical study described here show that the phase-based local feature performs better than local differential invariants for common illumination changes

and 2-D rotation, while giving similar results for scale changes of up to 20%. We are currently investigating the use of brightness renormalization for the local differential invariants, as in [19], in order to reduce the brightness sensitivity of the differential invariant approach and provide a fairer comparison.

1.1 Previous Work

The use of local features is usually associated with the object recognition task. Currently, object recognition methods are of three types, namely: 1) systems that match geometric features, 2) systems that match luminance data, and 3) systems that match robustly detectable, informative, and relatively sparse local features. The first type of system, namely those that utilize geometric features (see [2, 6, 9, 12]), are successful in some restricted areas, but the need of user-input models makes the representation of some objects, such as paintings or jackets, extremely hard. View-based methods (see [11, 15, 22]) have avoided this problem since they are capable of learning the object appearance without a user-input model. However they suffer from difficulties such as: 1) illumination changes are hard to be dealt with; 2) pose and position dependence; and 3) partial occlusion and clutter can damage the system performance (but see [1, 11]).

The third type of object recognition method is based on local image descriptors extracted from robustly detectable image locations. Systems that are based on this method show promising results mainly because they solve most of the problems in the view-based methods, such as illumination changes, clutter, occlusion, and segmentation, while keeping most of their improvements in terms of flexibility and simplified model acquisition. Rao and Ballard [17] explore the use of local features for recognizing human faces. The authors use principal component analysis (PCA) to reduce the dimensionality of *localized natural image patches* at *multiple scales* rather than PCA of entire images at a single scale. In [16], Nelson presented a technique to automatically extract a geometric description of an object by detecting semi-invariants at localized points. A new concept was presented by Schmid and Mohr [20], where, instead of using geometric features, the authors use a set of differential invariants extracted from interest points. In [13, 14] Lowe presents a novel method based on local scale-invariant features detected at interest points.

2 Image Deformations Studied

The image deformations considered here are: a) uniform brightness changes, b) non-uniform local brightness variations, c) noise addition, d) scale changes, and e) rotation changes. The uniform brightness change is simulated by adding a constant to the brightness value taking into account the non-linearity of the brightness visual perception, as follows:

$$I_h(\boldsymbol{x}) = 255 * \left[\max\left(0, \left(\frac{I(\boldsymbol{x})}{255} \right)^\lambda + k \right) \right]^{\frac{1}{\lambda}}, \tag{1}$$

where $\lambda = 2.2$, and k is the constant the alters the final brightness value. The resulting image is linearly mapped to values between 0 and 255, and then quantized.
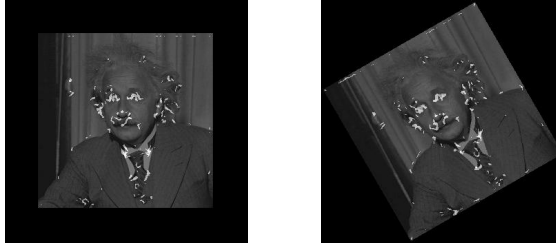
Fig. 1. Typical interest points detected on an image (brighter spots on the image). The right image shows the original points and the left one depicts the interest points detected after a 30°-degree rotation.

For the non-uniform local brightness variations, a highlight at a specific location of the image is simulated by adding a Gaussian blob in the following way:

$$I_h(\boldsymbol{x}) = I(\boldsymbol{x}) + 255 * G(\boldsymbol{x} - \boldsymbol{x}_0; \sigma), \tag{2}$$

where $\sigma = 10$, $\boldsymbol{x}_0$ is a specific position in the image, and $G(\boldsymbol{x}; \sigma) = \exp\left(-x^2/(2\sigma^2)\right)$. Again, the resulting image is mapped to values between 0 and 255, and then quantized.

For noise deformations, we simply add Gaussian noise with varying standard deviation ($\sigma = 255 * [10^{-3}, 10^{-1}]$), followed by normalization and quantization, as above. The last two deformations involve spatial image warps. In particular, we consider 2D rotations (from 0° to 180° in intervals of 30°) and uniform scale changes (with expansion factors in the range $[0.63, 1.58]$). Every image used in these deformation experiments is blurred, down-sampled and mapped to values between 0 and 255 in order to reduce high frequency artifacts caused by noise.

3 Interest Points

In the literature, view-based recognition from local information always relies on interest points, which represent specific places in an image that carry distinctive features of the object being studied. For example, in [13], interest points are represented by local extrema, with respect to both image location and scale, in the responses of difference of filters. Alternatively, a detector that uses the auto-correlation function in order to determine locations where the signal changes in two directions is used in [20]. A symmetry based operator is utilized in [10] to detect local interest points for the problem of scene and landmark recognition. In [16], a contour detection is run on the image, and points of high curvature around the shape are selected as interest points.

Here we consider the Harris corner detector (see [7]) used in [20], where a matrix that averages the first derivatives of the signal in a window is built as follows:

$$\mathbf{C}(\boldsymbol{x}) = \exp -\frac{x^2 + y^2}{2\sigma^2} \otimes \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}, \tag{3}$$

Fig. 2. The four images used for testing interest point detection. The right three images are also used for evaluating the local feature vectors.

where $\sigma = 2.0$, and $\otimes$ is the convolution operation. Here $I_x = G_x \otimes I$, where G_x is the x-derivative of a Gaussian with standard deviation 1, and similarly for I_y. The eigenvectors of this matrix encodes edge directions, while the eigenvalues, $\lambda_1(\boldsymbol{x})$ and $\lambda_2(\boldsymbol{x})$, represent edge strength. Corners, or interest points, can then be defined as locations at which $\lambda_1(\boldsymbol{x}) \geq \lambda_2(\boldsymbol{x}) \geq t$, where t is a threshold. Given the fact that the threshold function described in [7] does not produce a value between 0 and 1, we have found the following function to provide a more convenient threshold criterion:

$$R(\boldsymbol{x}) = \frac{\lambda_2(\boldsymbol{x})}{c + (1/2) * (\lambda_1(\boldsymbol{x}) + \lambda_2(\boldsymbol{x}))}, \tag{4}$$

where c is set based on the histogram of $R(\boldsymbol{x})$ of various types of images. Here, we select $c = 1$, and every point that has $R(\boldsymbol{x}) \geq 0.5$ is considered an interest point. Fig. 1 shows the corners detected for the Einstein image.

Two measures are computed to assess the performance of the interest point detector, namely the true positive rate and reliability. Given a point $\boldsymbol{x}_i$ in the original image space, and an image deformation specified by a matrix M and a translation vector $\boldsymbol{b}$, the transformed image location is

$$\boldsymbol{x_j} = M\boldsymbol{x_i} + \boldsymbol{b}. \tag{5}$$

Let us consider the set of interest points detected in an image I_i:

$$In(I_i) = \{x_i | R(\boldsymbol{x_i}) \geq 0.5\}, \tag{6}$$

where $\boldsymbol{x_i} \in I_i$.

The true positive (TP_{rate}) rate of interest point detection, between the original image I_i and the transformed image I_j, is based on the following measure:

$$TP_{rate} = \frac{|\{\boldsymbol{x_i} | \exists \boldsymbol{x_j} \, s.t. \, ||M\boldsymbol{x_i} + \boldsymbol{b} - \boldsymbol{x_j}|| < \epsilon\}|}{|In(I_i)|}. \tag{7}$$

where $x_i \in In(I_i)$, $x_j \in In(I_j)$, $|| \cdot ||$ denotes the *Euclidean* norm, and $\epsilon = 1.5$ pixels. However, this measure does not account for extraneous interest points in the transformed image. Therefore, we also measure the reliability of L_{rate} by calculating:
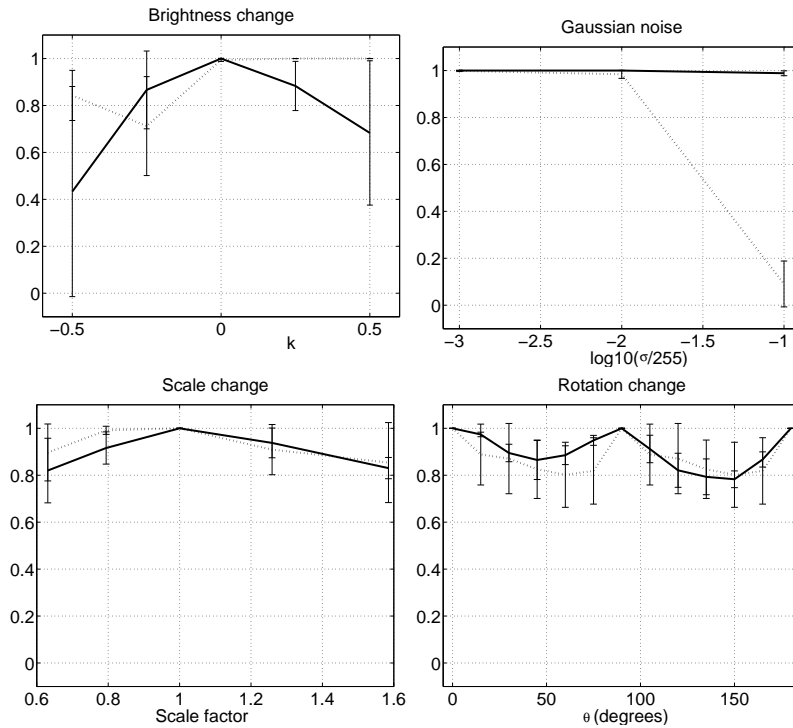
$$L_{rate} = \frac{|In(I_i)|}{|In(I_j)|}, \tag{8}$$

Fig. 3. Interest point repeatability. The graphs show the true positive (solid line) and reliability (dotted line) rates for the four types of image distortion.

where the maximum value for L_{rate} is constrained to be 1.

In Fig. 3, we can see common type of image deformations, the true positive rate and the reliability rate for the interest point detector (note: from left to right, image number 2 is extracted from the COIL database [18], and image number 4 is extracted from [21]).

4   Local Feature Vector

Ideally, the local features used to describe an object should have the following two properties: a) be complex enough to provide a strong information about a specific location of an image; and b) be relatively stable to changes in the object configuration, so that small transformations do not affect the efficiency of the correlation process. In this section we consider the problem of finding good candidates for such local features.

A recent proposal for local features described in [20] uses a set of derivatives, coined the "Local-Jet", that is invariant to rotation and is defined as follows:

$$\mathcal{V} = \begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ji} \\ \epsilon_{ij}(L_{jkl} L_i L_k L_l - L_{jkk} L_i L_l L_l) \\ L_{iij} L_j L_k L_k - L_{ijk} L_i L_j L_k \\ -\epsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix} \in \Re^9, \tag{9}$$

where we use the tensor summation convention, ϵ_{ij} is the 2-D anti-symmetric epsilon tensor defined by $\epsilon_{12} = -\epsilon_{21} = 1$ and $\epsilon_{11} = \epsilon_{22} = 0$, and $L_i = \frac{\partial}{\partial x_i} G(\boldsymbol{x}, \sigma) * I$ is the element of the local jet such that $G(\boldsymbol{x}, \sigma)$ is a *Gaussian* function, and I is the image.

Alternatively, in [13], after detecting interest points, the image is locally characterized by a set of Scale Invariant Feature Transform (SIFT) features that represents a vector of local image measurements.

4.1 Phase and Amplitude Information

We use a local feature approach, similar to the ones described above, but with a new type of feature using phase information. The phase-based local feature is a complex representation of local image data that is obtained through the use of quadrature pair filters, tuned to a specific orientation θ and scale σ. More specifically, we use the steerable quadrature filter pairs described in [5] as follows: Let

$$\begin{aligned} g(\boldsymbol{x}, \sigma, \theta) &= G_2(\sigma, \theta) * I(\boldsymbol{x}), \\ h(\boldsymbol{x}, \sigma, \theta) &= H_2(\sigma, \theta) * I(\boldsymbol{x}), \end{aligned} \tag{10}$$

where $G_2(\sigma, \theta)$ is the second derivative of a Gaussian, $H_2(\sigma, \theta)$ is the approximation of Hilbert transform of G_2, and σ is the standard deviation of the Gaussian kernel used to derive G_2 and H_2. A complex polar representation can be written as:

$$g(\boldsymbol{x}, \sigma, \theta) + ih(\boldsymbol{x}, \sigma, \theta) = \rho(\boldsymbol{x}, \sigma, \theta) e^{i\phi(\boldsymbol{x}, \sigma, \theta)}, \tag{11}$$

where $\rho(\boldsymbol{x}, \sigma, \theta)$ is the local amplitude information and $\phi(\boldsymbol{x}, \sigma, \theta)$ is the local phase information.

4.2 Saturating the Amplitude Information

The amplitude saturation is similar to contrast normalization (see [8]) and to the constraint on a minimum absolute amplitude (see [4]). It is desirable to allow the amplitude to saturate in order to reduce the system's sensitivity to brightness change. Therefore, whenever the local amplitude is high enough the saturated amplitude should be roughly constant. Here we use

$$\tilde{\rho}(\boldsymbol{x}, \sigma, \theta) = 1 - e^{\frac{-\rho^2(\boldsymbol{x}, \sigma, \theta)}{2 * \sigma_\rho^2}}, \tag{12}$$

where $\sigma_\rho = 2.5$. As a result, $\tilde{\rho}$ is roughly 1 for ρ over $2\sigma_\rho$, and near 0 for small amplitudes.

4.3 Local Image Description

Since a single pixel does not provide a distinctive response we consider several sample points, say $\{x_{i,m}\}_{m=1}^{M}$, taken from a region around each interest point, x_i. We use the sampling pattern depicted in Fig. 4, with the center point $x_{i,1}$ denoting the specific interest point x_i (the reasons for selecting this particular sampling pattern are discussed further below). At each spatial sample point $x_{i,m}$ the filters are steered to N equally spaced orientations, namely

$$\theta_n(x_i) = \theta_M(x_i) + (n-1)\frac{180^o}{N}, \quad \text{for } n = 1, \ldots, N. \tag{13}$$

Here $\theta_M(x_i)$ is the main orientation of the pixel computed as described in [5], except we use the sign of the imaginary response of the filter steered to this orientation to resolve a particular direction (i.e. mod 360°) from this orientation. Notice that this main orientation $\theta_M(x_i)$ therefore determines both the orientations that the filters are steered to and the positions of the sample points along circle centered on the interest point x_i (see Fig. 4).

The feature vector $F(x_i)$ has individual components specified by the saturated complex filter responses. We use $\tilde{\rho}_i(n, m)e^{\phi_i(n,m)}$ to denote the filter response evaluated at $x_{i,m}$ and steered to orientation $\theta_n(x_i)$, for $n = 1, \ldots, N$, and $m = 1, \ldots, M$. Together these responses form the NM-dimensional complex feature vector $F(x_i)$.

4.4 Phase Correlation

The similarity between local features is computed using phase correlation since this is known to provide some stability to typical image deformations such as brightness changes and near identity image warps. The similarity measure for our feature vector is the normalized phase correlation

$$S(F(x_i), F(x_j)) = \left| \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} \tilde{\rho}_i(n, m)\tilde{\rho}_j(n, m)e^{i(\phi_i(n,m) - \phi_j(n,m))}}{1 + \sum_{m=1}^{M} \sum_{n=1}^{N} \tilde{\rho}_i(n, m)\tilde{\rho}_j(n, m)} \right|. \tag{14}$$

The reason for adding the 1 in the denominator above is to provide a low-amplitude cutoff for the normalization. This results in similarity values $S(F(x_i), F(x_j)) \in [0, 1]$.

4.5 Feature Vector Configuration

An empirical study was conducted to select the remaining parameters of the local feature vector. These are: a) the number of steering directions, N; b) the number of sample points $P = M - 1$ on the circle surrounding the interest point x_i; and c) the radius, l, of the circle. Each of these parameters represents a compromise between stability (better for small values of the parameters), and expressiveness (better at larger values). By evaluating the detection rates and false target rates (in the manner described for the experiments below) we selected $M = 9$, $N = 4$, and $l = 3$ as providing a reasonable trade-off between expressiveness and stability.
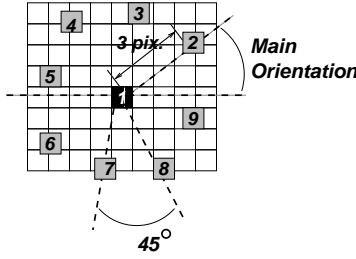
Fig. 4. Configuration of local descriptor.

5 Experiment Setup

In order to compare our feature with the differential invariant feature in (9), 3 test images were selected (see Fig. 2), and 5 image databases (see Fig. 5) were selected consisting of 12 images each. None of the test images were included in these databases. Some of the images inserted into the databases were selected due to appearance similarities with the test images, and other images were just natural scenes.

Given the 5 types of image deformations studied (see Section 2), the comparison is based on the Receiver Operating Characteristics (ROC) curves where the detection rate vs false positive rate is computed for each of the local feature types. In order to define these rates, let x_i be an interest point in a test image. Suppose $x_i^0 = M x_i + b$ denotes the correct position of this interest point in the transformed test image, according to the spatial deformation used. The detection rate (DT) is then defined to be the proportion of interest points x_i such that there exists some interest point, x_j in the transformed image which is both sufficiently close to the mapped point (i.e. $||x_j - x_i^0|| < \epsilon$) and which has a similar local feature vector (i.e. $S(F(x_i), F(x_j)) > \tau$). Here ϵ was fixed at 1.5 pixels, while τ was varied to generate the ROC curves. Similarly, given this same interest point x_i in the test image, a false positive is defined by the presence of a similar interest point x_j in the database (i.e. $S(F(x_i), F(x_j)) > \tau$). The false positive rate (FP) is defined to be the number of these false positives divided by the total number of test image interest points evaluated.

The threshold for both similarity functions is varied as follows: for the phase correlation, that has values in $[0, 1]$, the variation step is 0.1; the differential invariant feature uses the Mahalanobis distance, as described in [20], which can have practically any value above 0, so the variation step is 1 until $DT_{rate} \geq 0.99$. The actual curves are computed using intervals of 0.03 for the false positive rate, and these are plotted using linear interpolation.

6 Results

Fig. 6 shows the ROC curve for uniform brightness changes. It is clear that the phase-based feature displays consistently better results, and, due to amplitude saturation, the feature is almost unaffected by an increase in brightness. However, it is more sensitive
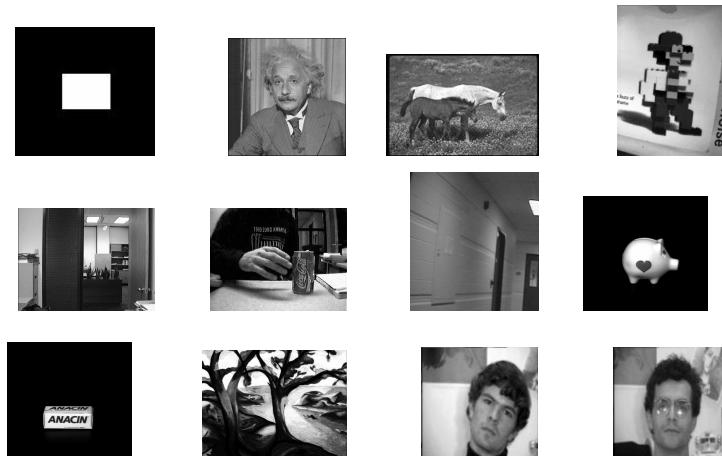
Fig. 5. Database of images.

to decreases in brightness, which is presumably due to the appearance of unsaturated low amplitude responses. The differential invariant feature, on the other hand, is seen to be quite sensitive to these changes. This is also clear from Fig. 7, where we show the detection rate for thresholds τ at which the false positive rate is fixed at 0.1. It is clear from this plot that the phase-based approach is much less sensitive to brightness changes. The same is true for non-uniform brightness changes, as shown in Fig. 8.

The phase-based feature also gives good results for other types of image deformations. As shown in Fig. 9, the performance of both types of features is seen to be similar for additive Gaussian noise. For scale changes, the differential invariant feature is seen to have a somewhat larger detection rate, for the same level of false positives(see Figure 10). This is true primarily for the larger scale changes. For scale changes between $\pm 20\%$, the phase-based local feature provides comparable performance (see Fig. 11). Finally, the ROC curves show the phase-based feature is somewhat better under image rotation (see Fig. 12). In order to control for small brightness changes which may have occurred during the rotation and scale deformations, we computed the ROC curves with and without rescaling the transformed image to values between 0 and 255. Both cases gave similar results to the ones reported here.

7   Conclusions and Future Work

A new type of local feature based on the phase and amplitude of steerable bandpass filters is proposed here. An empirical study is conducted in order to demonstrate that it has the basic characteristics necessary for useful local features, that is, they are robust to common image deformations and distinctive. Moreover, an empirical comparison with differential invariant features shows that the phase-based local feature performs better in terms of common illumination changes and 2-D rotation, while giving comparable or
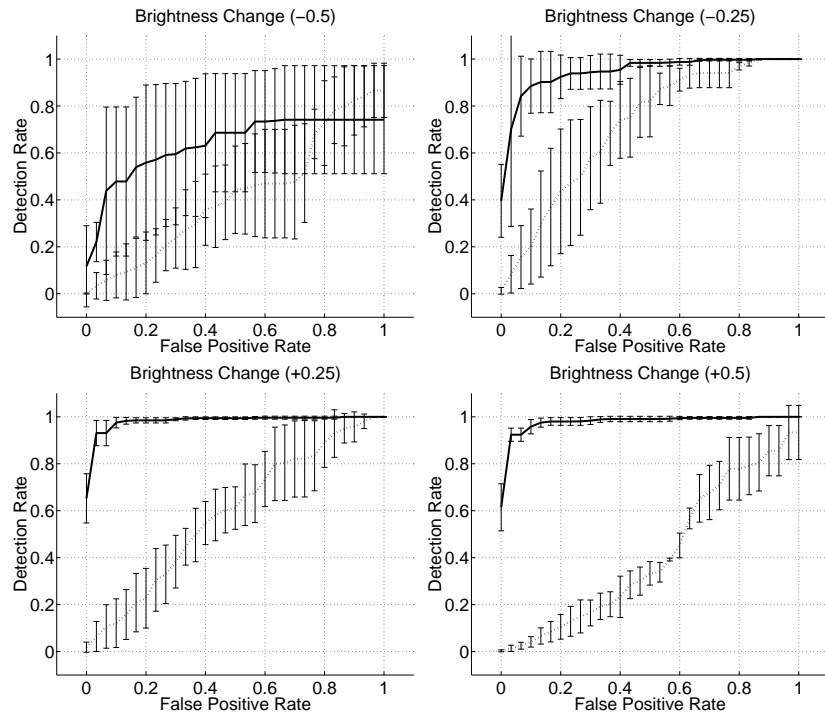
Fig. 6. Uniform brightness changes. Solid line represents the phase-based feature. Dotted line represents differential invariant feature.
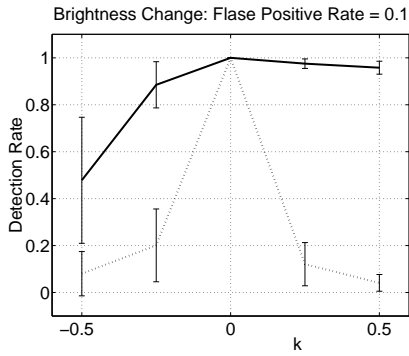

Fig. 7. Uniform brightness changes with a false positive rate fixed at 0.1 and computing the detection rate for varying amount of change. Solid line represents the phase-based feature. Dotted line represents differential invariant feature.
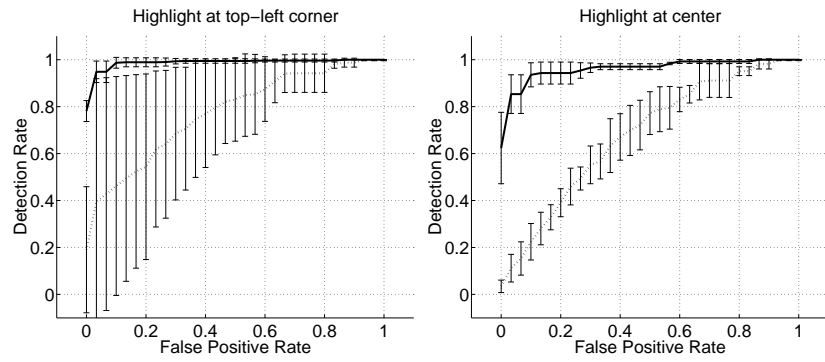
Fig. 8. Non-uniform local brightness changes. Solid line represents the phase-based feature. Dotted line represents differential invariant feature.
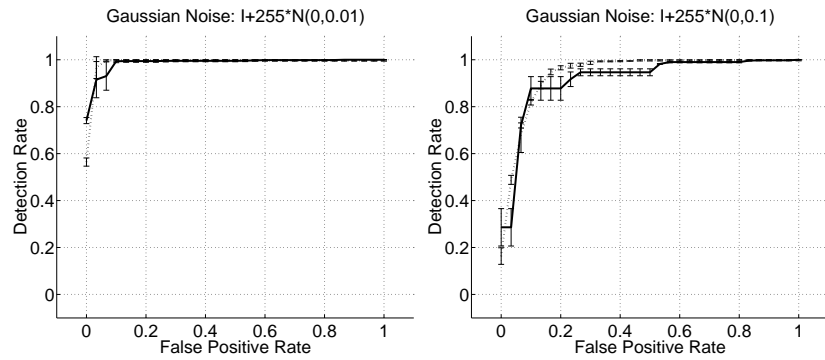

Fig. 9. Gaussian noise changes. Solid line represents the phase-based feature. Dotted line represents differential invariant feature.
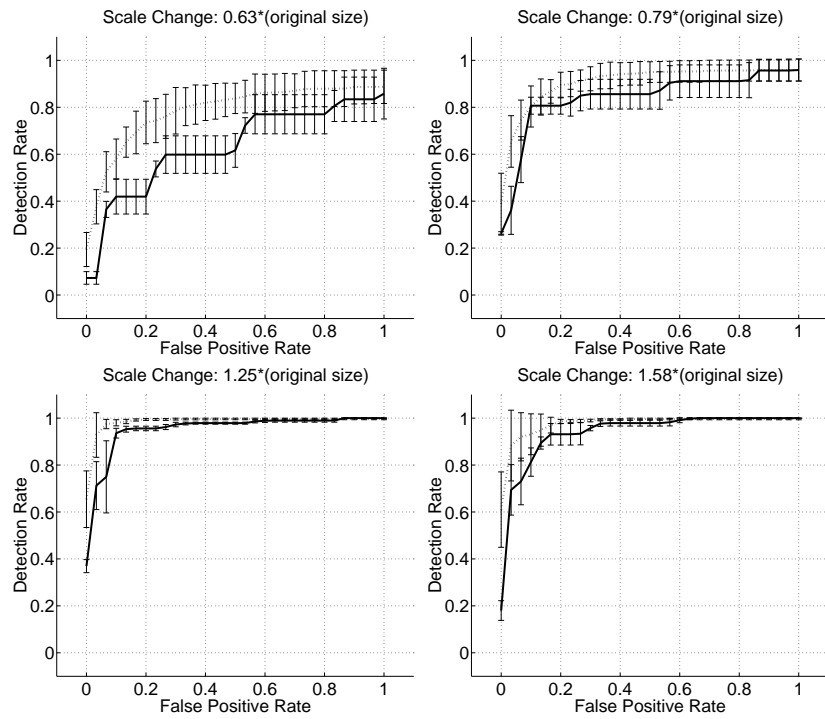
Fig. 10. Scale changes. Solid line represents the phase-based feature. Dotted line represents differential invariant feature.
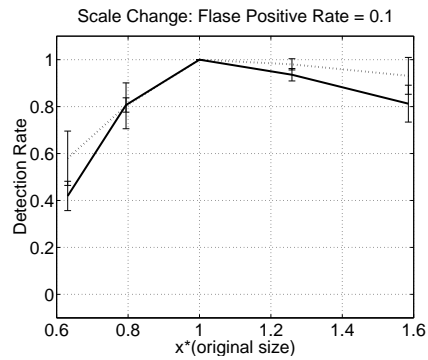

Fig. 11. Scale changes with a false positive rate fixed at 0.1 and computing the detection rate for varying amount of change. Solid line represents the phase-based feature. Dotted line represents differential invariant feature.
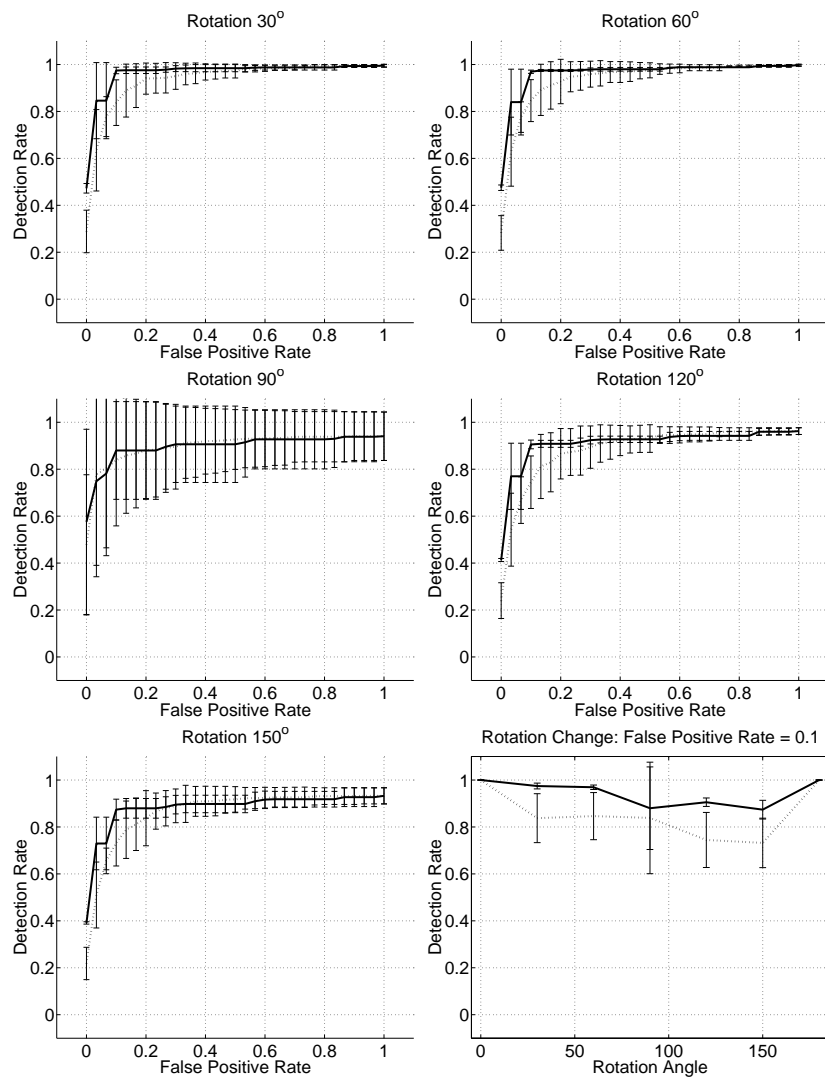
Fig. 12. Rotation changes. Solid line represents the phase-based feature. Dotted line represents differential invariant feature. The graph at the bottom-right corner shows rotation changes with a false positive rate fixed at 0.1 and computing the detection rate for varying amount of change.

slightly worse results when dealing with scale changes. An important area for further study is the use of brightness normalization in the differential invariant features, and the comparison of the result with our phase-based approach.

The phase-based local feature has obvious applications in object recognition, but a few issues must be dealt with before exploring its capabilities. The interest point detector used here can be replaced by another one that provides a better response in terms of the information being studied here, (i.e., phase and amplitude), and presents a better response to scale changes. The high dimensionality of the vector can represent a burden, so methods for reducing it, or search methods that perform well in high dimensions must be sought. Finally, grouping features before searching the database of models is an important component that should be added.

References

1. M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. In *4th European Conf. on Computer Vision*, pages 329–342, Cambridge, April 1996.
2. R. Brooks. Model-based 3-d interpretations of 2-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):140–150, 1983.
3. D. Fleet. *Measurement of Image Velocity*. Kluwer Academic Publishers, 1992.
4. D. Fleet, A. D. Jepson, and M. Jenkin. Phase-based disparity measure. In *CVGIP: Image Understanding*, pages 198–210, 1991.
5. W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.
6. W. E. L. Grimson and T. Lozano-Pérez. Localizing overlapping parts by searching the interpretation tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):469–482, 1987.
7. C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.
8. D. J. Heeger. Computational model of cat striate physiology. Technical report, Massachusetts Institute of Technology, October 1989.
9. D. Huttenlocher and S. Ullman. Object recognition using alignment. In *International Conference on Computer Vision*, pages 102–111, London, UK, 1987.
10. D. Jugessur and G. Dudek. Local appearance for robust object recognition. In *IEEE Computer Vision and Pattern Recognition*, pages 834–839, Hilton Head, USA, June 2000.
11. A. Leonardis and H. Bischoff. Dealing with occlusions in the eigenspace approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 453–458, San Francisco, USA, June 1996.
12. D. G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3):355–395, 1987.
13. D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, Corfu, Greece, September 1999.
14. D. G. Lowe. Towards a computational model for object recognition in it cortex. In *First IEEE International Workshop on Biologically Motivated Computer Vision*, pages 20–31, Seoul, Korea, May 2000.
15. H. Murase and S. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.
16. R. C. Nelson. Memory-based recognition for 3-d objects. In *ARPA Image Understanding Workshop*, pages 1305–1310, Palm Springs, USA, February 1996.

17. R.P.N. Rao and D.H. Ballard. Natural basis functions and topographic memory for face recognition. In *International Joint Conference on Artificial Intelligence*, pages 10–17, 1995.

18. S. K. Nayar S. A. Nene and H. Murase. Columbia object image library (coil-20). Technical report, Department of Computer Science, Columbia University, February 1996.

19. B. Schiele and J.L. Crowley. Object recognition using multidimensional receptive field histograms. In *4th European Conference on Computer Vision*, volume 1, pages 610–619, April 1996.

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

21. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

22. M. Turk and A. P. Pentland. Face recognition using eigenfaces. In *IEEE Computer Vision and Pattern Recognition*, pages 586–591, 1991.