Research Article

Feature Point Extraction from the Local Frequency Map of an Image

Jesmin Khan,¹ Sharif Bhuiyan,¹ and Reza Adhami²

¹Department of Electrical Engineering, Tuskegee University, Tuskegee, AL 36088, USA

²Department of Electrical and Computer Engineering, University of Alabama in Huntsville, Huntsville, AL 35899, USA

Correspondence should be addressed to Jesmin Khan, khanj@mytu.tuskegee.edu

Received 18 March 2011; Accepted 12 December 2011

Academic Editor: Tamal Bose

Copyright © 2012 Jesmin Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a novel technique for detecting rotation- and scale-invariant interest points from the local frequency representation of an image. Local or instantaneous frequency is the spatial derivative of the local phase, where the local phase of any signal can be found from its Hilbert transform. Local frequency estimation can detect edge, ridge, corner, and texture information at the same time, and it shows high values at those dominant features of an image. For each pixel, we select an appropriate width of the window for computing the derivative of the phase. In order to select the width of the window for any given pixel, we make use of the measure of the extent to which the phases, in the neighborhood of that pixel, are in the same direction. The local frequency map, thus obtained, is then thresholded by employing a global thresholding approach to detect the interest or feature points. Repeatability rate, a performance evaluation criterion for an interest point detector, is used to check the geometric stability of the proposed method under different transformations. We present simulation results of the detection of feature points from image utilizing the suggested technique and compare the proposed method with five existing approaches that yield good results. The results prove the efficacy of the proposed feature point detection algorithm. Moreover, in terms of repeatability rate; the results show that the performance of the proposed method with respect to different aspect is compatible with the existing methods.

1. Introduction

Interest point detection is a recent terminology in computer vision that refers to the detection of interest points for subsequent processing. An interest point is a point in an image which has a well-defined position and can be robustly detected. This means that an interest point can be a corner but it can also be, for example, an isolated point of local intensity maximum or minimum, line endings, or a point on a curve where the curvature is locally maximal. In practice, most of the so-called corner detection methods detect interest points in general rather than corners in particular. In the literature, “corner,” “interest point,” and “feature” are used somewhat interchangeably. An interest point is a point in the image which in general can be characterized as follows:

- (i) it has a clear, preferably mathematically well-founded, definition;
- (ii) it has a well-defined position in image space;

- (iii) the local image structure around the interest point is rich in terms of local information contents, such that the use of interest points simplify further processing in the vision system;
- (iv) it is stable under local and global perturbations in the image domain, including deformations as those arising from perspective transformations (sometimes reduced to affine transformations, scale changes, rotations, and/or translations) as well as illumination/brightness variations, such that the interest points can be reliably computed with high degree of reproducibility.

Historically, the notion of interest points goes back to the earlier notion of corner detection, where corner features were in early work detected with the primary goal of obtaining robust, stable, and well-defined image features for object tracking and recognition of objects from two-dimensional images. In practice, however, most corner detectors are

sensitive not specifically to corners, but to local image regions which have a high degree of variation in all directions. Today, a main application of interest points is to signal points/regions in the image domain that are likely candidates to be useful for image matching, object recognition, motion detection, tracking, image mosaicing, panorama stitching, 3D modeling, and so forth.

There are a wide variety of methods reported in the literature for interest point and corner detection in grey-level images. However, many true corner points are missed and false interest points are detected in an image. Current detection methods can be categorized into three types: contour-based, parametric-model-based, and intensity-based methods. Contour-based methods first extract contours and then search for maximal curvature or inflexion points along the contour chains or carry out some polygonal approximation and then search for intersection points. Parametric model methods fit a parametric intensity model to the signal. They often provide subpixel accuracy, but are limited to specific types of interest points, for example, to L-corners. Intensity-based methods compute a measure that indicates the presence of an interest point directly from the grey values. This type of detector does not depend on edge detection or mathematical models.

Our paper presents a novel affine invariant interest point detector. The main contribution of this work is the extraction of interest points from local frequency map of an image. The proposed technique cannot be explicitly placed into any of the three categories. But it can be considered as a contour-based method, because local frequency estimation detects edges of an image. For the computation of local frequency, we use local phase which is a local neighborhood property. Therefore, in order for the value of that property to be useful, it must be computed in a neighborhood, which requires the determination of the appropriate size of the window that defines the local structure. This requirement is the same as the problem of scale selection. In this paper we introduce a method of scale selection which is based on a newly defined local image property called *phase divergence*. Local phase of any 1D signal can be calculated by utilizing a quadrature filter called the Hilbert transformer. In our work we have defined an equivalent 2D Hilbert transformer for the computation of local phase of an image. By definition, the spatial derivative of local phase is called the local or instantaneous frequency. Finally, a global thresholding technique is applied to the local frequency image to extract interest points.

The requirement for good feature point detectors is that they should be invariant to image transformations and correspond to the same regions of an image for different viewpoints. That means, with the change in the shape, size, and orientation, the detector automatically adapts based on the underlying image intensities. In particular, considering images from different geometric transformations, the interest points detected after the transformations should be the same so that the transformed versions of the points detected in the original image and points detected after the transformation of the image commute. They have often been called invariant interest points in the literature, though in principle they change covariantly with the transformation.

The fact is that, even though the regions themselves are covariant, the normalized image pattern they cover and the feature descriptors derived from them are typically invariant. The proposed method is largely invariant to significant affine transformation including large rotations, shearing, and scale changes. Such transformations introduce significant changes in point locations as well as in the scale and the shape of the neighborhoods of interest points. Our approach addresses these problems simultaneously and offers invariance to geometric transformation.

In this paper we evaluate the proposed method utilizing the “repeatability” [1] criteria, which directly measures the quality of the detected feature points for tasks such as image matching, object recognition, and 3D reconstruction. It is complementary to localization accuracy, which is relevant for tasks such as camera calibration and 3D reconstruction of specific scene points. Repeatability and localization are conflicting criteria: smoothing improves repeatability but degrades localization [2]. Repeatability explicitly compares the geometrical stability of the detected interest points between different images of a given scene taken under varying viewing conditions. An interest point is “repeated” if the 3D scene point detected in the first image is also accurately detected in the second transformed image. The repeatability rate is the percentage of the total observed points that are detected in both images. Our detector satisfies the criteria well. The proposed detector is compared to five existing methods which have been shown to yield good results. Utilizing repeatability the proposed method is shown to yield comparable to improved performance.

The remainder of the paper is organized as follows. Section 2 presents a state-of-the-art on some existing interest point detectors. The motivation behind this work has been described in Section 3. The proposed interest point detection algorithm is given in Section 4. Experimental results along with the performance evaluation employing repeatability rate are demonstrated in Section 5. The comparison with five other existing methods is also given in this section. Concluding remarks and recommendations for future improvement are given in Section 6.

2. Related Work on Interest Point Detection

In this section we briefly present some existing interest points detection methods for each of the three categories.

2.1. Contour-Based Methods. Contour-based methods have existed for some long time. Bandera et al. [3] and Rattarangsi and Chin [4] utilize chain codes or curvature to detect boundary corners of an object. Rattarangsi and Chin use boundary curvature as an input function for further Gaussian scale-space analysis to detect corners. Bandera et al. estimate the curvature of a contour in an adaptive fashion so that corners can be detected even if they appear at different natural scales. Horaud et al. [5] extract line segments from image contours that are grouped, and intersections of grouped line segments are utilized as interest points. Shilat et al. [6] first detect ridges and troughs in the images. High

curvature points along ridges or troughs, or intersection points, are utilized as interest points. Mokhtarian and Suomela [7] describe an interest point detector based on two sets of interest points. One set consists of T-junctions extracted from edge intersections, and the second set is obtained using a multiscale framework. The two sets are compared, and close interest points are merged. Pikaz and Dinstein [8] propose an algorithm based on a decomposition of noisy digital curves into a minimal number of convex and concave sections. They use the properties of pairs of sections that are determined in an adaptive manner to detect interest points.

Besides the above-mentioned contour-based interest points detection schemes, we specially want to give particular emphasis to interest point detection methods based on wavelet transform because wavelet transforms (WTs) may be considered as a form of time-frequency representation, and local or instantaneous frequency is the term or parameter that accounts for the time-varying nature of the spectral characteristics (in particular the frequency of the spectral peaks) of nonstationary signal.

In the past decade, interest point detection schemes using the WT became popular as a result of the fact that the WT is able to decompose an input signal into smooth and detailed parts by low-pass and high-pass filters at multiresolution levels [9]. In this manner, local deviations can be easily captured at various detailed decomposition levels. Several wavelet-based approaches [10–12] directly utilized the slope between two boundary points (e.g., tangent orientation) on an object contour as input and analyzed the wavelet modulus sum utilizing an interlevel ratio between multiresolution levels. These approaches detect corner candidates (including true corners and circular arcs) in a first step and eliminate circular arcs in a second by predefined thresholds. For instances, Quddus and Gabbouj [13] utilize optimal scale and pre-defined threshold to localize true corners precisely for a complicated shape (e.g., a shark). Hua and Liao [14] use x and y parametric coordinates to replace curvatures and tangent orientations as an input to WT analysis. This approach achieves a higher identification rate for polygon-shaped objects. Loupias et al. [15, 16] present a salient point detector that extracts points where variations occur in the image, whether they are corner like or not. The detector is based on WT for the detection of global and local variations. After reviewing wavelet-based researches, Yeh [17] recently concluded that 1D WT is a robust corner detection scheme due to its excellent local deviation capturing capability. In their work they transform the 2D boundaries of an object into a 1D $\theta - P$ representation (where θ is the tangent angle variations of the arc length, P along the object's boundary) and they detect the feature points utilizing the 1D wavelet coefficients at the finest (first) detailed decomposition level.

2.2. Intensity-Based Methods. Moravec [18] develops one of the first signal-based interest point detectors. His detector is based on the autocorrelation function of the signal. Beaudet's [19] detector uses the second derivatives of the signal to compute the measure $DET = I_{xx}I_{yy} - I_{xy}^2$, where $I(x, y)$

is the intensity surface of the image. Points where this measure is maximal are defined as interest points. Kitchen and Rosenfeld [20] define an interest point detector that uses the curvature of planar curves.

Several interest point detectors [21–24] are based on a matrix related to the autocorrelation function. The utilized matrix, A , averages derivatives of the signal in a window W around a point (x, y) :

$$A(x, y) = \begin{bmatrix} \sum_W (I_x(x_k, y_k))^2 & \sum_W I_x(x_k, y_k) I_y(x_k, y_k) \\ \sum_W I_x(x_k, y_k) I_y(x_k, y_k) & \sum_W (I_y(x_k, y_k))^2 \end{bmatrix}, \quad (1)$$

where $I(x, y)$ is the image function and (x_k, y_k) are the points in the window W around (x, y) . This matrix captures the structure of the neighborhood. If this matrix is of rank two, an interest point is detected. A matrix of rank one indicates an edge and a matrix of rank zero a homogeneous region.

Heitger et al. [25] develop an approach that extracts 1D directional characteristics by convolving the image with orientation-selective Gabor-like filters. The first and second derivatives of the 1D characteristics are computed in order to obtain 2D characteristics. Smith and Brady [26] compare the brightness of each pixel in a circular mask to the center pixel to define an area, and two-dimensional features are detected from the size, centroid, and second moment of this area. The approach proposed by Laganieri [27] is based on a variant of the morphological closing operator which uses two closing operators and four structuring elements. The first closing operator is sensitive to vertical/horizontal L-corners and the second to diagonal L-corners. Lowe [28] presents a method for extracting distinctive invariant features from images by taking a cascade filtering approach, which has four major stages: (1) identifying potential interest points that are invariant to scale and orientation by searching across all possible scales, (2) fitting a detailed model at each candidate location to select keypoints based on measures of their stability, (3) assigning one or more orientations to each keypoint based on local image gradient directions, and (4) measuring the local image gradients at the selected scale around each keypoint to allow for significant local shape distortion and change in illumination.

2.3. Parametric-Model-Based Methods. The parametric model used by Rohr [29] is an analytic junction model convolved with a Gaussian kernel. Rohr uses an interest point detector which maximizes $DET(A)$ of (1) for A as well as the intersection of line segments to determine the initial values for the model parameters. Deriche and Blaszkia [30] develop an acceleration of Rohr's method. They substitute an exponential for the Gaussian smoothing function. Baker et al. [31] propose an algorithm that represents a feature as a densely sampled parametric manifold in a low-dimensional subspace. A feature is detected if the projection of the surrounding intensity values in the subspace lies sufficiently close to the feature manifold. Parida et al. [32] use a deformable template to detect radial partitions. The

minimum description length principle determines the optimal number of partitions that best describes the signal.

3. Motivation

Local frequency representation is a useful preprocessing operation for a number of image processing and image exploitation tasks. The local frequency extracted from a given image provides a very useful and robust feature for execution of tasks that involve comparison of image frames for classification purposes and/or combining distinct frames into a single better quality image [33]. In particular, image registration and image fusion are specific tasks that can be executed with greater precision by employing local frequency representations obtained from the component image frames to be registered. Recent research has shown that registration algorithms that implement interest point matching ideas can be tailored to deliver improved performance (more accurate estimation of registration parameters, registration of dissimilar imagery data captured by sensors operating in different modalities, etc.) when the interest points used for matching are obtained from local frequency representations [34].

The following properties of local frequency make it a valuable candidate for an invariant image representation: (1) local frequency estimate, unlike gradient, is relatively invariant to signal energy, in other words, also independent to the image contrast. This makes a notable advantage over ordinary gradient-based edge detector; for example, in some images there are some important features that reside in a region with a quiet small contrast against background and gradient methods will probably fail to detect them efficiently, while local frequency representation can do this work well; (2) local frequency estimation can detect the structure of the scene in the image (high frequency features, that is, edge, ridge, and texture information at the same time). This imposes an advantage over the methods, where the edge and ridge have to be extracted by different operators; (3) the third notable advantage that makes local frequency representations valuable is that it is relatively invariant to illumination changes and has a good localization to the original signal in the spatial domain (thus enjoying structure-preserving properties). Based on the above facts, in this paper local frequency map has been suggested to be employed as an efficient approach for extracting interest points.

Despite the usefulness of these representations, there do not exist computationally efficient procedures for extraction of local frequencies from digital imagery data, especially for large-format images. Given a digital image, computation of local frequency typically involves evaluating the gradient of the phase of the analytic form of the image [33], which in turn is obtained by processing the image with a bank of analytic filters, typically a set of 2D Gabor filters characterized by a set of parameters (spatial frequency, orientation angle, and the filter bandwidth), and each tuned to a specific frequency value or to a range of frequencies. Tuning this set of Gabor filters that can cover the entire frequency space for the given image is critical for obtaining

an accurate representation of the local frequency; however, choosing the size of the filter bank is often done in an ad hoc manner and one usually errs on the conservative side by employing a larger than needed set of filters in order to cover the entire range of frequencies. In this paper, we propose a computationally efficient approach for developing the local frequency representation from a given image. In doing so, we formulate a synonymous 2D interpretation of the quadrature filter (Hilbert transformer). Moreover, this new algorithm provides an interesting solution for choosing the size of the quadrature filter adaptively for each pixel based on the image content and hereby ensures the generation of the precise value of the local frequency corresponding to the image intensity at each pixel of an image.

4. Proposed Algorithm

The outline of the interest point extraction algorithm presented in this paper is illustrated in Figure 1 that involves three stages. (1) The first step of the proposed approach for interest point extraction is to formulate a quadrature filter called the Hilbert transformer with customized size, $w \times w$ for each pixel. We consider 8 different sizes of the filter as $w \times w$, for $w = 2k - 1$ and $k = 1, 2, \dots, 8$. In this process we need to compute local phase for each of the filter window size $w = 2k - 1$ at pixel (x, y) , which is represented as $\phi_k(x, y)$. We make use of an image attribute named phase divergence, $\nabla \phi_k$ to select the appropriate size of the filter $\hat{w}_k(x, y) \times \hat{w}_k(x, y)$. (2) The next step is to compute the local phase $\hat{\phi}_k(x, y)$ for the pixel (x, y) at the selected scale $\hat{k}(x, y)$. Thereafter, the local frequency $\hat{f}_k(x, y)$ for pixel (x, y) at the selected filter size is estimated by taking the spatial derivative of the local phase. (3) At the final step, a global thresholding is applied to the local frequency image to extract the interest points. Local frequency representation of an image efficiently detects all the important features, even those that reside in the regions with a quiet small contrast against background. Accordingly, at the true feature points, the local frequency layout of an image shows distinctly higher values than at the points which carry less, or no, information. Thus, points where the frequency measure of the image is greater than a pre-defined threshold are set as interest points. The following subsections discuss each step of the algorithm in detail.

4.1. Local Phase Measurement. In order to describe the local phase of a real signal, we will first introduce the concept of analytic signal in one-dimension (1D) which can be generalized to higher dimensions [33]. Given a signal $g(t)$ in 1D, its analytic signal is defined as $g_A(t) = g(t) - jg_H(t)$, where $g_H(t)$ is the Hilbert transform of $g(t)$. $g_H(t)$ is defined as $g_H(t) = (1/\pi t) * g(t) = (1/\pi) \int_{-\infty}^{\infty} (g(\tau)/(\tau - t)) d\tau$, where $*$ denotes 1D linear convolution. For a 1D discrete signal, the Hilbert transform is the convolution of the signal with a discrete filter, h_1 , of size $1 \times w$, where

$$h_1(n) = \begin{cases} 1 & \text{if } n = 0 \\ \frac{1}{\pi n} & \text{if } n \neq 0, n \in \left\{ \frac{w-1}{2}, \frac{w+1}{2} \right\} \end{cases} \quad (2)$$

FIGURE 1: Block diagram of the proposed algorithm.

Although the above discussion is given for a one-dimensional signal for the sake of simplicity, it can be generalized to higher dimensions readily. For image-processing applications, evaluation of the local frequency for a given image can be performed using 2D functions.

In two-dimensional (2D), the Hilbert transform is not uniquely defined and the extension of this transform into 2D is not well known in the area of digital image processing [35]. Hence, we have devised an equivalent representation of the 2D version of the quadrature filter (Hilbert transformer) for the computation of local phase of an image. As we mentioned, the 2D discrete quadrature filter, H_2 , of size $w \times w$ is defined as, $H_2 = h'_1 * h_1$, where $'$ and $*$, respectively, denote matrix transpose and matrix multiplication.

To get the Hilbert transform of discretely sampled image data, we convolve the image $I(x, y)$ with this filter. For the image $I(x, y)$, the corresponding analytic image $I_A(x, y)$ can be defined as $I_A(x, y) = I(x, y) - jI_H(x, y)$, where $I_H(x, y)$ is

the Hilbert transformation of $I(x, y)$. The Hilbert transform, $I_H(x, y)$, is defined by $I_H(x, y) = H_2 * I(x, y)$, where $*$ denotes 2D linear convolution. Thus, the transformation of a given image $I(x, y)$ to the corresponding analytic image $I_A(x, y)$ can be regarded as the result of convolving the real image with a complex filter, such as the Hilbert transformer. The argument of $I_A(x, y)$ is referred to as the local phase of $I(x, y)$, which is defined in the spatial domain. For the window size $w = 2k - 1$, we represent the phase at pixel (x, y) by $\phi_k(x, y)$.

4.2. Scale Selection. We have proposed a new image property called *phase divergence* that is used for selecting the controlling size of the filter window, which is also called the integration scale [21, 36]. Phase divergence is a measure of the extent to which the phase angles in a certain window around a pixel are in the same direction with respect to

the phase angle of that pixel. There are other methods that have already been utilized in the literature for scale selection; for example, in [37, 38] the image property called *polarity* for selecting scale is used. In that paper, the polarity is computed in the neighborhood of a given pixel with respect to the dominant orientation of that pixel. The scale selection method in [36] is based on localizing the peak across scale of the determinant or trace of the second moment matrix computed for each pixel within a window. The phase divergence at pixel (x, y) for a given window size $w = 2k - 1$, as we define it, is

$$\nabla\phi_k(x, y) = \frac{|\phi_{k+}(x, y) - \phi_{k-}(x, y)|}{\phi_{k+}(x, y) + \phi_{k-}(x, y)}, \quad (3)$$

where $\phi_{k+}(x, y)$ is the number of phase angles in the window w , that are on the positive side of the phase angle at pixel (x, y) and $\phi_{k-}(x, y)$ is the number of phase angles in the window, w , that are on the negative side of the phase angle at pixel (x, y) .

The phase divergence, $\nabla\phi_k$, ranges from 0 to 1 and it varies as the size of the window, w , changes. By computing the phase divergence at all pixels of the image for $w = 2k - 1$, $k = 1, 2, \dots, 8$, we produce a stack of phase divergence images across scale, k . Then, the phase divergence image at scale k is smoothed by convolving with a Gaussian filter of size, w . Finally the scale is selected based on the derivative of the phase divergence with respect to scale. For a given pixel (x, y) , the scale $\hat{k}(x, y)$ is selected as the first value of $k(x, y)$ for which the difference between values of phase divergence at successive scales ($\nabla\phi_k(x, y) - \nabla\phi_{k-1}(x, y)$) is less than 1 percent. In the uniform regions of an image, the selected scale should be 1, because uniform region appears not to change across scale. A region can be declared to be uniform if the mean contrast (standard deviation of intensity) of that region across scale is less than 0.1 [37].

4.3. Frequency Selection. Local frequency can be estimated by using quadrature filters [33, 34]. Given a real signal, its corresponding analytic signal is complex with the real part being the original signal itself and the imaginary part being its Hilbert transform obtained by convolving the signal with a quadrature filter. As discussed before, the argument of the analytic signal is referred to as the local phase of the original signal. The spatial derivative of the local phase gives the local frequency of an image. For the scale k we denote the frequency at pixel (x, y) as $f_k(x, y)$, which is estimated by computing the gradient of the phase values in the window of size w .

The gradient image is simply the derivative of the local image values. Mathematically, the gradient of a two-variable function (here the image intensity function) is a 2D vector at each image point with the components given by the derivatives in the horizontal and vertical directions. At each image point, the gradient vector points in the direction of largest possible intensity increase, and the length of the gradient vector corresponds to the rate of change in that direction.

To estimate local frequency $f_k(x, y)$, we use the most common method of computing the gradient of the local phase image which is the simple local derivative

$$\begin{aligned} \text{grad}(\phi_k(x, y)) &= \left(\frac{\partial\phi_k(x, y)}{\partial x}, \frac{\partial\phi_k(x, y)}{\partial y} \right) \\ &= (\nabla\phi_{kx}(x, y), \nabla\phi_{ky}(x, y)) \\ f_k(x, y) &= \sqrt{\nabla\phi_{kx}^2(x, y) + \nabla\phi_{ky}^2(x, y)}, \end{aligned} \quad (4)$$

where $\nabla\phi_{kx}(x, y) = \partial\phi_k(x, y)/\partial x = \phi_k(x+1, y) - \phi_k(x, y)$ and $\nabla\phi_{ky}(x, y) = \partial\phi_k(x, y)/\partial y = \phi_k(x, y+1) - \phi_k(x, y)$ are the numerical gradients (differences) in the x and y directions, respectively.

The local frequency images are computed for $w = 2k - 1$, $k = 1, 2, \dots, 8$, and thus there are a stack of local frequency images across scale, k . After selecting the appropriate scale $\hat{k}(x, y)$ for a given pixel (x, y) , we compute frequency $\hat{f}_k(x, y)$ at that scale, and finally we have the accurate value of the local frequency at each pixel of the image.

4.4. Interest Point Determination. The local frequency image representation is found to be stable to describe and capture all the meaningful and salient features of an image. For an image, which is a kind of nonstationary signal, the local frequency is an important characteristic; it is a time-varying parameter which defines the location of the signals spectral peak as it varies with time. Conceptually, at the edges, ridges and corners (they are the dominant features of an image and associated with high-frequency components), the local frequency representation has high values.

As an example for a one-dimensional signal the local frequency estimation is demonstrated in Figure 2. Figure 2(a) is the original one-dimensional signal. It contains both edges and ridges with different contrasts. Figure 2(b) shows the final local frequency representation at the selected scale $\hat{k}(n)$ customized for each point n of that 1D signal.

To evaluate the performance resulting from the present approach for extracting local frequency representation from an image data and to establish its robustness to scene details, we conducted experiments on images, and the results are presented in Figure 3. For the reference image shown in Figure 3(a), Figure 3(b) shows the local frequency representation at the selected scale $\hat{k}(x, y)$ customized for each point (x, y) of that 2D signal in Figure 3(a). As we have seen for 1D signal, once again the local frequency representation given in Figure 3(b) clearly displays the edges, ridges, and corners which are extracted perfectly by the algorithm. These results serve to illustrate the versatility of the approach used here to process 1D or 2D signals and the robustness, accuracy, and utility of the algorithm to extract local frequency information.

Three noticeable characteristics are demonstrated in this experiment: (1) the local frequency representation yields outputs at the edges, ridges, and corners of a signal; (2) the local frequency, unlike gradient, is relatively independent of signal contrast; (3) the localization of the detector is quiet faithful to the original signal. The latter property is due

FIGURE 2: (a) A 1D signal and (b) local frequency map.

FIGURE 3: (a) A 2D image and (b) local frequency map.

to the stability of phase information [9]. The scale space properties of local frequency estimation, as shown in Figures 2 and 3, have long been exploited in the computer vision community both from a computational and localization of feature point of view. The scale-space, that is, the multiresolution properties of local frequency estimation, leads to computationally efficient algorithms and alleviates the problem of local minima entrapment when used in the context of interest point detection problem as described below.

For the selection of the interest or feature points, the local frequency image, $\hat{f}_k(x, y)$, is thresholded. A global non-histogram-based thresholding technique has been incorporated rather than local (adaptive) thresholding in this step [39]. The threshold level, γ , is determined by $\gamma = \sum(\hat{f}_k(x, y) \cdot s) / \sum s$, where “ $\cdot$ ” denotes pixel wise multiplication and s is given by $s = \max(|g_1 * \hat{f}_k(x, y)|, |g_2 * \hat{f}_k(x, y)|)$, where $g_1 = [-101]$, $g_2 = [-101]^T$, and “ $*$ ” denotes two-dimensional linear convolution. The binary image $I_{bf}(x, y)$ is then given by

$$I_{bf}(x, y) = \begin{cases} 1 & \text{if } \hat{f}_k(x, y) > \gamma, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

A thinning algorithm is applied to this binary image yielding $I_f(x, y)$, which contains the extracted interest points. The algorithm for finding true interest points makes three passes through the thinned local frequency image. First, points are selected if the value of the frequency exceeds a

threshold determined globally. Second, the selected points that are not locally maximum in the original intensity image in its 3×3 neighborhood are deleted. In the final pass, the subset of pixels are kept such that the minimum distance between any pair of points is larger than a given threshold.

Consider an image of size $M \times N = n_1$ pixels such that $F_1 = \{p_i = (x_i, y_i); i = 1, 2, 3 \dots n_1\}$ contains all the pixel coordinates. Let F_2 be the first selection of the feature points from the local frequency image, which forms the set $F_2 \subset F_1$

$$F_2 = \left\{ p_i : I_f(x_i, y_i) = 1 \right\} \\ = \{p_i = (x_i, y_i); i = 1, 2, 3, \dots, n_2 : n_2 < n_1\}, \quad (6)$$

where n_1 is the number of all pixels and n_2 is the number of selected points after discarding unwanted points by thresholding.

At the next two passes, from the points in F_2 we retain those points that are locally maximum in their $W_m = 3 \times 3$ neighborhood with the restriction that the distance between any two feature points is larger than a given threshold (this is set to 5 pixels in our experiment). Thus, $F_3 \subset F_2$ is the set of feature points that are locally maximum in the frequency image, $\hat{f}_k(x, y)$, and $F_f \subset F_3$ is the final set of feature points after discarding those closely spaced points

$$F_3 = \left\{ p_i = (x_i, y_i) : \hat{f}_k(x_i, y_i) = \max_{\{j \in i-W_m/2, i+W_m/2\}} \hat{f}_k(x_j, y_j) \right\} \\ F_f = \{p_i = (x_i, y_i); i = 1, 2, 3, \dots, n_3 : n_3 < n_2 < n_1\}, \quad (7)$$

where n_3 is the number of selected points after discarding redundant points from F_2

$$F_f = \{p_i : |p_i - p_{i-1}| > 5 \text{ pix}, |p_i - p_{i+1}| > 5 \text{ pix}\},$$

$$F_f = \{p_i = (x_i, y_i); i = 1, 2, 3, \dots, n_f : n_f < n_3 < n_2 < n_1\}, \quad (8)$$

where n_f is the number of selected points after discarding redundant points from F_3 .

Following the above procedure, we select the interest points from an image.

5. Experimental Results

In this section, we provide the results of experiments which are conducted to test the efficacy of the proposed interest point detection algorithm. In order to test the immunity of our algorithm to transformation, the original images are scaled, rotated, and sheared. We evaluate the proposed detector using the repeatability rate for image rotation and scale change. Besides these, the repeatability rates of five interest point detectors are compared with our method under different image rotation and scale changes.

For comparison we have chosen five detectors such that those detectors have already been reported to offer good performance. Among the five chosen detectors Harris', Lowe's, and Tomasi's methods are intensity-based methods. These are chosen because Harris' method has been reported to be better than any other detector, Lowe's algorithm also known as scale-invariant feature transform (SIFT) is the best scale invariant detector and Tomasi's detector is the best for tracking applications. The other two detectors are chosen because (1) those are contour-based methods like our method and (2) those methods use wavelet decomposition.

Simulation results of the five methods on real image are presented in Figure 4. For the reference image shown in Figure 3(a), feature points extracted by Harris' approach, Lowe's procedure, Tomasi' method, Loupias' technique, and Yeh's algorithm are shown in Figures 4(a), 4(b), 4(c), 4(d), and 4(e), respectively. The points detected by the presented method are given in Figure 4(f). From this figure it can be seen that points selected by the proposed method cover all the curvatures of object boundaries and yield the true corner points.

For the evaluation of rotation invariance of the detector, Figures 5 and 6 show the detection results for two rotated versions of the reference image. Figures 5(f) and 6(f) give the results of our method, where the rotation angle for the images in Figure 5 is 40° and for the images in Figure 6 is 110° . The performance of Harris', Lowe's, Tomasi's, Loupias', and Yeh's methods are given in Figures 5(a), 5(b), 5(c), 5(d), and 5(e), respectively, for the rotation 40° and in Figures 6(a), 6(b), 6(c), 6(d), and 6(e), respectively, for the rotation 110° . From the figure, it can be said that Harris's, Lowe's, and the proposed methods give the best result for both rotations. The performance of Tomasi's technique is better than Loupias' and Yeh's methods.

The effect of image scale change on detection result is tested and demonstrated in Figures 7 and 8. The detected

points by the proposed technique are shown in Figures 7(f) and 8(f), where the scale change for Figure 7 is 1.5 and for Figure 8 is 3.4. Points detected by Harris', Lowe's, Tomasi', Loupias' and Yeh's methods are presented in Figures 7(a), 7(b), 7(c), 7(d), and 7(e), respectively, for the scale change 1.5 and in Figures 8(a), 8(b), 8(c), 8(d), and 8(e) respectively for the scale change 3.4. It can be seen from the figure that all the methods are scale invariant.

To evaluate the functioning of all of the five detectors as affine-invariant systems, nonuniform scaling is applied in some directions to have shearing in the reference image shown in Figure 3(a). Figure 9(f) displays the feature points detected in the sheared image by the proposed method. The results of detection by Harris', Lowe's, Tomasi's, Loupias', and Yeh's methods are presented in Figures 9(a), 9(b), 9(c), 9(d), and 9(e), respectively. By examining the figure it can be said that the proposed method performs satisfactorily for detecting interest points from the sheared image as well.

From the figures it can be concluded that the proposed method can be used as an affine- and scale-invariant detector. To complement the subjective evaluations, we present a qualitative performance evaluation of the proposed affine-invariant detector. The stability and accuracy of the detectors are evaluated using the repeatability criterion [1]. The repeatability score for a given pair of images is the average number of corresponding points detected in those images under different geometric and photometric transformations. We take into account only the points located in the part of the scene present in both images. The repeatability rate is measured for a localization error of 1.5 pixels; that is the detected point lies in a very close pixel neighborhood of the predicted point. Measuring the repeatability rate within 1.5 pixels or less, the probability that two points are accidentally within the error distance is negligible.

We first compare the detectors for image rotation followed by scale change. The repeatability rate as a function of the angle of image rotation is displayed in Figure 10(a). The rotation angles vary between 0° and 180° . Under repeatability, Lowe's methods give the best results for all rotations, where this algorithm obtains a repeatability rate of about 78%. From the observation of the plot of the repeatability rate with image rotation, the proposed technique does not outperform the SIFT algorithm, but it yields the same performance as Harris' method with a repeatability rate of about 75% for all rotations. Notably, it offers better performance than Tomasi's, Loupias', and Yeh's techniques, where these approaches offer a repeatability rate of about 60%, 45%, and 50%, respectively.

Figure 10(b) shows the repeatability rate as a function of scale changes. The scale varies from 1 to 4 for the plot of repeatability rate as a function of scale. The results show that all the detectors are scale sensitive except Lowe's method. As the name implies, the SIFT algorithm proposed by Lowe offers the best performance with scale change. This method is the least dependent on the change in scale. Harris', Tomasi's, and the proposed detectors give reasonable results, with the repeatability rate as a decreasing function of scale change. Loupias' and Yeh's methods are very sensitive to scale change, and the results of these methods are hardly usable.

FIGURE 4: Feature points detected by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupias' technique, (e) Yeh's algorithm, and (f) proposed technique.

FIGURE 5: Feature points detected in 40° rotated image by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupias' technique, (e) Yeh's algorithm, and (f) proposed technique.

FIGURE 6: Feature points detected in 110° rotated image by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupias' technique, (e) Yeh's algorithm, and (f) proposed technique.

FIGURE 7: Feature points detected in 1.5 times scaled image by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupias' technique, (e) Yeh's algorithm, and (f) proposed technique.

FIGURE 8: Feature points detected in 3.4 times scaled image by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupias' technique, (e) Yeh's algorithm, and (f) proposed technique.

FIGURE 9: Feature points detected in sheared image by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupias' technique, (e) Yeh's algorithm, and (f) proposed technique.

FIGURE 10: Plot for the repeatability rate as a function of (a) rotation angle and (b) change in scale.

FIGURE 11: Feature points detected in the second image by (a) Harris' method, (b) Lowe's method, (c) Tomasi's method, (d) Loupious' technique, (e) Yeh's algorithm, and (f) proposed technique.

Above a scale factor of about 2, the results from all the detectors, except Lowe's method, are mainly due to artifacts, because at larger scales many more points are found in the textured regions of the scene, so accidental correspondences are more likely.

For the evaluation of detection performance, the feature points extracted by the proposed method and the five other algorithms are presented for one more image in Figure 11, where the detected points are superimposed on the original image to evaluate the interest points location. From this

FIGURE 12: (a) Interest points detected from the binary image of letter “h” by (b) the proposed algorithm and (c) the SIFT algorithm.

figure it is observed that the proposed method extract points where variations occur in the image, that is, where the image information is supposed to be the most important. Additionally, the set of detected interest points are not cluttered in few regions but rather spread out at different parts of the image. Most importantly, the proposed method covers all the curvatures of object boundaries and yields at the true features i.e., the edges, ridges, and corners. Accordingly, the extracted points detect the structure of the scene and lead to a complete image representation.

From the parallel observation of the figures of detection results and the value of calculated repeatability rate, it can be claimed that the proposed method compares favorably against other well-known methods. Based on the plot for the repeatability rate, like Harris’ and Lowe’s algorithms, the proposed technique is minimally dependent on the rotations, which is a desirable and attractive characteristic for any feature point detector. Though the proposed technique does not outperform the SIFT algorithm, on average, it offers performance similar to Harris’ method and it yields better results than Tomasi’s, Loupious’ and Yeh’s techniques. Notably, it offers better performance than the other two contour-based methods: Loupious’ and Yeh’s approaches.

Therefore, as a contour-based technique, the presented approach can be expected to perform well for the applications where true corner points are needed to be detected from the edges, ridges, and corners for further processing. As an example, for a binary image, interest points should not be found in the uniform region of constant intensity, that is, either in the background or foreground. They would rather lie only on the edges. As shown in Figure 12, for the binary image of the letter “h,” the SIFT algorithm detects feature points in the uniform region. But the proposed method, as a contour-based technique, extracts interest points only from the edges, which signifies the performance variation of different algorithms depending on the types of image and/or applications. From the simulation results, it can be concluded that the proposed method can be utilized as a dependable affine-invariant interest point detector.

In addition to the accuracy, the speed, in other words, the computational cost of the proposed algorithm is measured. The interest point detection process takes 25 seconds for

an image of size 100×100 pixels. The processing time is recorded based on the simulation of MATLAB code on a 2.59 GHz Pentium IV machine. For the same image, the computation time for Harris’, Lowe’s, Tomasi’s, Loupious’ and Yeh’s methods is 0.8 seconds, 1.8 seconds, 2.2 seconds, 3.6 seconds, and 1.4 seconds, respectively. We are working on making the proposed interest point detection scheme faster by parallel processing of an image with the employment of digital signal processor and C++ programming language.

6. Conclusion

This research presents a robust, rotation-invariant, and scale-invariant corner detection scheme on image based on the local or instantaneous frequency representation of an image. The image is convolved with the Hilbert transformer customized for each pixel to calculate the local phase image. We have formulated a scale selection technique based on the image property phase divergence, which is used to select the controlling size of the filter window. The phase image is differentiated to obtain the local frequency representation of the image. By the employment of a global thresholding to the frequency image, the proposed technique localizes all the true corner points on edges, ridges, and corners. Under appropriate image resolution and region of support, the proposed approach precisely captures the true corner points and is free from the false alarms on circular arcs for both simple and complicated objects in varying rotation and scale conditions.

The proposed technique detects feature points globally from the whole image at a time based on the value of the frequency at a given pixel, and the detected interest points are largely independent of the imaging conditions; therefore, points are geometrically stable. This approach has strong robustness of detecting good feature points, as it detects feature points from the regions with even a quiet small contrast against background. Experimental results also suggest that the proposed local-frequency-based corner detection approach is stable and efficient. The proposed method is a generic concept and can find its application in many matching and recognition problems.

In future work, we want to tackle a number of challenges for further improvement in the detection of interest points. In this respect, we are planning to address the following issues: (1) to improve the repeatability rate with both rotation and scaling, (2) to demonstrate the application of the proposed interest point detector for object recognition, and (3) to reduce the computation cost by parallel processing and implementing the algorithm in C++.

References

- [1] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.
- [2] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [3] A. Bandera, C. Urdiales, F. Arrebola, and F. Sandoval, "Corner detection by means of adaptively estimated curvature function," *Electronics Letters*, vol. 36, no. 2, pp. 124–126, 2000.
- [4] A. Rattarangsi and R. T. Chin, "Scale-based detection of corners of planar curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 430–449, 1992.
- [5] R. Horaud, T. Skordas, and F. Veillon, "Finding geometric and relational structures in an image," in *Proceedings of the 1st European Conference on Computer Vision*, pp. 374–384, Antibes, France, 1990.
- [6] E. Shilat, M. Werman, and Y. Gdalyahu, "Ridge's corner detection and correspondence," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 976–981, Puerto Rico, Puerto Rico, USA, June 1997.
- [7] F. Mokhtarian and R. Suomela, "Robust image corner detection through curvature scale space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1376–1381, 1998.
- [8] A. Pikaz and I. H. Dinstein, "Using simple decomposition for smoothing and feature point detection of noisy digital curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 808–813, 1994.
- [9] A. Bruce and H. Y. Gao, *Applied Wavelet Analysis with SPLUS*, Springer, New York, NY, USA, 1996.
- [10] J. S. Lee, Y. N. Sun, and C. H. Chen, "Multiscale corner detection by using wavelet transform," *IEEE Transactions on Image Processing*, vol. 4, no. 1, pp. 100–104, 1995.
- [11] C. H. Chen, J. S. Lee, and Y. N. Sun, "Wavelet transformation for gray-level corner detection," *Pattern Recognition*, vol. 28, no. 6, pp. 853–861, 1995.
- [12] A. Quddus and M. M. Fahmy, "Fast wavelet-based corner detection technique," *Electronics Letters*, vol. 35, no. 4, pp. 287–288, 1999.
- [13] A. Quddus and M. Gabbouj, "Wavelet-based corner detection technique using optimal scale," *Pattern Recognition Letters*, vol. 23, no. 1–3, pp. 215–220, 2002.
- [14] J. P. Hua and Q. M. Liao, "Wavelet-based multi scale corner detection," in *Proceedings of the 5th International Conference on Signal Processing*, pp. 341–344, 2000.
- [15] E. Loupias, N. Sebe, S. Bres, and J. M. Jolion, "Wavelet-based salient points for image retrieval," in *Proceedings of the IEEE Conference on Image Processing (ICIP '00)*, pp. 518–521, September 2000.
- [16] Q. Tian, N. Sebe, M. S. Lew, E. Loupias, and T. S. Huang, "Image retrieval using wavelet-based salient points," *Journal of Electronic Imaging*, vol. 10, no. 4, pp. 835–849, 2001.
- [17] C. H. Yeh, "Wavelet-based corner detection using eigenvectors of covariance matrices," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2797–2806, 2003.
- [18] H. P. Moravec, "Towards automatic visual obstacle avoidance," in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, p. 584, Cambridge, Mass, USA, 1977.
- [19] P. R. Beaudet, "Rotationally invariant image operators," in *Proceedings of the 4th International Joint Conference on Pattern Recognition*, pp. 579–583, Tokyo, Japan, 1978.
- [20] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," *Pattern Recognition Letters*, vol. 1, no. 2, pp. 95–102, 1982.
- [21] W. Forstner, "A framework for low level feature extraction," in *Proceedings of the 3rd European Conference on Computer Vision*, pp. 383–394, Stockholm, Sweden, 1994.
- [22] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Alvey Vision Conference*, pp. 147–151, 1988.
- [23] C. Tomasi and T. Kanade, "Detection and tracking of point features," Tech. Rep. CMU-CS-91-132, Carnegie Mellon University, 1991.
- [24] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 593–600, Seattle, Wash, USA, June 1994.
- [25] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kubler, "Simulation of neutral contour mechanisms: from simple to end-stopped cells," *Vision Research*, vol. 32, no. 5, pp. 963–981, 1992.
- [26] S. M. Smith and J. M. Brady, "SUSAN—a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.
- [27] R. Laganier, "Morphological corner detection," in *Proceedings of the 6th IEEE International Conference on Computer Vision*, pp. 280–285, Bombay, India, January 1998.
- [28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [29] K. Rohr, "Recognizing corners by fitting parametric models," *International Journal of Computer Vision*, vol. 9, no. 3, pp. 213–230, 1992.
- [30] R. Deriche and T. Blaszk, "Recovering and characterizing image features using an efficient model based approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 530–535, New York, NY, USA, June 1993.
- [31] S. Baker, S. K. Nayar, and H. Murase, "Parametric feature detection," *International Journal of Computer Vision*, vol. 27, no. 1, pp. 27–50, 1998.
- [32] L. Parida, D. Geiger, and R. Hummel, "Junctions: detection, classification, and reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 687–698, 1998.
- [33] G. Granlund and H. Knutsson, *Signal Processing for Computer Vision*, Kluwer Academic, Dodrecht, The Netherlands, 1995.
- [34] J. Liu, "Regularized quadrature filters for local frequency estimation: application to multimodal volume image registration," in *Proceedings of the Conference on Vision Modeling and Visualization*, pp. 507–514, November 2001.
- [35] M. R. Arnison, C. J. Cogswell, N. I. Smith, P. W. Fekete, and K. G. Larkin, "Using the Hilbert transform for 3D

visualization of interference contrast microscope images,” *Journal of Microscopy*, vol. 199, no. 1, pp. 79–84, 2000.

- [36] J. Gårding and T. Lindeberg, “Direct computation of shape cues using scale-adapted spatial derivative operators,” *International Journal of Computer Vision*, vol. 17, no. 2, pp. 163–191, 1996.
- [37] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blob-world: image segmentation using expectation-maximization and its application to image querying,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, 2002.
- [38] T. Leung and J. Malik, “Detecting, localizing and grouping repeated scene elements from an image,” in *Proceedings of the European Conference on Computer Vision*, pp. 546–555, 1996.
- [39] S. U. Lee, S. Yoon Chung, and R. H. Park, “A comparative performance study of several global thresholding techniques for segmentation,” *Computer Vision, Graphics and Image Processing*, vol. 52, no. 2, pp. 171–190, 1990.