Acta Automatica Sinica
Vol. 40, No. 10 October, 2014
Available online at www.sciencedirect.com
ScienceDirect

A Scale Invariant Interest Point Detector in Gabor Based Energy Space

CAO Zheng-Cai[1, 2] MA Feng-Le[1, 2] FU Yi-Li[2] ZHANG Jian[3]

Abstract Interest point detection is a fundamental issue in many intermediate level vision problems and plays a significant role in vision systems. The previous interest point detectors are designed to detect some special image structures such as corners, junctions, line terminations and so on. These detectors based on some simplified 2D feature models will not work for image features that differ significantly from the models. In this paper, a scale invariant interest point detector, which is appropriate for most types of image features, is proposed based on an iterative method in the Gabor based energy space. It detects interest points by noting that there are some similarities in the phase domain for all types of image features, which are obtained by different detectors respectively. Firstly, this approach obtains the positions of candidate points by detecting the local maxima of a series of energy maps constructed by Gabor filter responses. Secondly, an iterative algorithm is adopted to select the corresponding characteristic scales and accurately locate the interest points simultaneously in the Gabor based energy space. Finally, in order to improve the real-time performance of the approach, a fast implementation of Gabor function is used to accelerate the process of energy space construction. Experiments show that this approach has a broader applicability than the other detectors and has a good performance under rotation and some other image changes.

Key words Interest point detection, Gabor filter, energy map, scale invariant

Citation Cao Zheng-Cai, Ma Feng-Le, Fu Yi-Li, Zhang Jian. A scale invariant interest point detector in Gabor based energy space. *Acta Automatica Sinica*, 2014, **40**(10): 2356−2363

Local features have been proved to be particularly appropriate to image matching, object recognition and 3D reconstruction as well as to many other computer vision applications since they are robust to occlusion, background changes and more distinctive than global features. Especially in a mobile vision system, which is required to deal with partial occlusion and other local disturbances, the introduction of local features is quite favorable and promising. In these typical applications, a common approach is to use interest point detectors to estimate a reduced set of local image regions. Then the interest point descriptors, which define these regions by their spatial locations, orientations, scales and possibly affine transformations, are built to find reliable image-to-image or image-to-model matches. So making the interest points be stable to various kinds of image transformations is a primary issue.

Over the past few years, many approaches have been proposed to detect interest points based on some special 2D image features, such as corners, blobs, multi-junctions and so on. Among them, the Harris detector was proposed for corner structures and has been proved to be the best one in robustness under rotation, illumination, variation and image noise[1]. On this basis, the Harris-Laplacian algorithm could detect corner-like features with the characteristic scales obtained by the Laplacian measures, which makes the points invariant to scale and limited viewpoint change[2]. The Hessian-Laplacian detector is a similar method with a little difference in expression of the second moment matrix. After analyzing the affine regions around the interest points, the detector was improved to deal with affine change which contains scale and viewpoint change. This detector is referred to as Harris-Affine detector[3].

Lindeberg extracted different types of image structures such as blobs, corners, ridges and edges by using different expressions of Gaussian derivatives, and then searched for 3D maxima of the scale normalized Laplacian-of-Gaussian (LoG) function for automatic scale selection. Then Lowe used the Difference-of-Gaussian (DoG) filters to approximate the LoG method to build scale space and proposed an efficient algorithm which is sensitive to blob-like features and has great success in image registration and object recognition[4]. Furthermore, the SURF detector combined the integral image and the box type filters to replace the DoG filter and achieved good performance in settling real time problems[5]. Besides the corner-like and blob-like features, the edge is also the common structures for interest point detection[6]. It is always assumed that the edge belongs to 1D features, so some other operations on the early extracted edges are taken for building some special criteria. Recently a low level image invariant called phase congruency have been proposed to represent the image features[7−8] and extended to corners or other feature detection[9−10]. The phase congruency based methods also have been used in many image applications[11−12]. No matter how many feature points these methods detect, the models they are based on fail to make them detect and localize all types of 2D features. This disadvantage limits the applications of the detectors and makes them easily influenced by the change of image contents, such as the change of background, and moving one object into the camera view.

Different from these model based feature detectors, this paper proposes an interest point detector which focuses on the similarity that all these features are related to points which contain maximal 2D order in the phase domain. So the detector is not limited by the types of models and is proved to be appropriate for all the types of features mentioned before. In real applications, one simple type of feature is usually difficult to meet the need of a high level task. In recent years many algorithms based on multi-feature fusion have been proposed in object recognitions, image

Manuscript received September 9, 2013; accepted April 23, 2014
Supported by National Natural Science Foundation of China (61105089), State Key Laboratory of Robotics and System (SKLRS-2013-ZD-03), and Open Foundation of the State Key Laboratory of Fluid Power Transmission and Control (GZKF-201212)
Recommended by Associate Editor ZHOU Jie
1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China 2. State Key Laboratory of Robotics and Systems, Harbin Institute of Technology, Harbin 150001, China 3. School of Mechanical Engineering, Tongji University, Shanghai 200092, China

tracking and retrieval. For examples, a method based on weighted fusion of color, texture and edge was used to resist interference and occlusion in video tracking; a weighted distance of various features was treated as the criterion in content based image retrieval. This is partly because adoption of many types of features can overcome the defect and limited condition of a single feature. Meanwhile there is another problem that the different standards of judgment and normalization in different detectors are difficult in fusion. However, the algorithm, which could detect different types of features simultaneously, has the internal connection with the features and has a comparatively advantage in deciding the weight factor in multi-feature fusion. The detector in this paper is a good choice to meet the requirement above.

The detected points are invariant to image translation, scale, rotation, illumination or limited viewpoint change. Comprehensive experiments show its adaptability for most of the 2D features and the better performance than some other approaches mentioned before.

1 Multi-scale interest point detector

The multi-scale interest points contain two-part information: positions in the spatial domain and characteristic scales which represent the scope of image structures. Therefore, the solution of interest point detection consists of two parts: the position detection and the characteristic scale selection.

In Subsection 1.1, the basic Gabor function is reviewed and compared. In Subsection 1.2, the energy based interest point detector is improved to multi-scale space. In Subsection 1.3, the Gabor based scale selection is explained. To reduce the complexity of the algorithm, a fast implementation of Gabor filter is described in Subsection 1.4.

1.1 The Gabor function

The Gabor functions, which have achieved good performances in many areas of image processing and pattern recognition, are bandpass filters with Gaussian modulated by complex sinusoids. Moreover, many recent neurophysiologic evidences from the visual cortex of mammalian brains suggest that the filter response profiles of the main class of linearly-responding cortical neurons are best modeled as a family of self-similar 2D Gabor wavelets. The basic 2D Gabor function is defined as follows:

$$g(x,y,\sigma,\theta) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\exp\left(\frac{\sigma i}{\pi}(x\cos\theta+y\sin\theta)\right) \quad (1)$$

where σ is the scale variation which denotes the width of the Gaussian envelop in space domain, and θ indicates the orientation of Gabor function. The 2D Gabor wavelet has been proved to be appropriate to energy concentration of the input image through multi-scale and multi-orientation, since it could achieve the theoretical limit for conjoint resolution of information in the spatial and Fourier domains, which is an advantage in signal analysis.

The Gabor functions, which have been widely used in facial, texture and object recognition, are designed as multi-scale and multi-directional filters to decompose the 2D image signal into a high dimensional space, where the characteristic of image signal is easily displayed and classified. After that, the Gabor features are represented by different kinds of descriptors or selected by some dimensionality reduction techniques such as principal component analysis (PCA) and linear discriminant analysis (LDA). Compared to these applications aimed at special image structures with fixed parameters, the Gabor functions used in this paper are expected to form a quadrature pair of local energy filters and construct an energy space. Then an iterative search is used to find the 3D maxima in the energy space, which represent the locations of the feature points and characteristic scales simultaneously. The iteration process with gradually adjusted parameters could be viewed as an automatic way to select prominent Gabor features compared with the previous methods and satisfies the condition of focus-of-attention (FoA).

1.2 Multi-scale interest points detection

A number of researches in biological vision show that the phase based features have high information content and play an important role in human perception compared with the usually defined image features such as edges, shadows and bars. The phase congruency, which is a low-level invariant property of image features, contains a wide range of features[7−8]. These features always occur at maxima of energy map which are proportional to the phase congruency maxima. Instead of treating the energy map as an intermediate variable to calculate the phase congruency, the proposed algorithm detects maxima of local energy map constructed by Gabor filter responses as the feature points.

Similar to the local energy of 1D signal, which is defined as convolution between the signal and a pair of orthogonal filters, the local energy of an image, for each orientation, is calculated as the square root of sum of squares of the image convolved with a quadrature pair of filters in the spatial domain.

$$g_{\sin} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\sin\left(\frac{\sigma}{\pi}(x\cos\theta+y\sin\theta)\right) \quad (2)$$

$$g_{\cos} = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\cos\left(\frac{\sigma}{\pi}(x\cos\theta+y\sin\theta)\right) \quad (3)$$

The odd symmetric filter (denoted as $g_{\sin}$) and even symmetric filter (denoted as $g_{\cos}$) obtained from the complete Gabor function have zero mean, identical L^2 norm, are orthogonal to each other and form a quadrature pair of energy filters to calculate 2D local energy map through the orientation θ.

Let W be the image filtered by a complex 2D Gabor filter; the even part and odd part are given respectively as follows:

$$W_{\rm even}(x,y,\sigma,\theta) = I(x,y)\otimes g_{\cos}(x,y,\sigma,\theta) \quad (4)$$

$$W_{\rm odd}(x,y,\sigma,\theta) = I(x,y)\otimes g_{\sin}(x,y,\sigma,\theta) \quad (5)$$

where $\otimes$ denotes the convolution symbol.

Then the oriented local energy map at the special scale σ and from orientation θ is denoted as $E(x,y,\sigma,\theta)$:

$$E(x,y,\sigma,\theta) = \|W(x,y,\sigma,\theta)\| \quad (6)$$

where $\|\cdot\|$ denotes modulus and is equal to the square root of the sum of the even and odd parts of W. The peak positions of energy map represent the significant features.

In order to detect features at all orientations, a bank of filters directed to different orientations is designed to ensure the energy map tile the frequency plane uniformly. It is noteworthy that the adding of energy maps obtained from all orientations with a same weight makes the features at different orientations be treated equally and invariant to rotation.