*ICIC Express Letters*
*Part B: Applications*
Volume **7**, Number **3**, March **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 525–530

# SCALE-INVARIANT INTEREST POINT DETECTION IN IMAGES BASED ON COMPLEX NETWORK ANALYSIS

Xuecheng Sun[1], Kun Zhao[1], Zheming Lu[2,*]
Tingyuan Nie[1] and Lijian Zhou[1]

[1]School of Communication and Electronic Engineering
Qingdao University of Technology
No. 11, Fushun Road, Qingdao 266033, P. R. China
sterling1982@163.com

[2]School of Aeronautics and Astronautics
Zhejiang University
No. 38, Zheda Road, Hangzhou 310027, P. R. China
[*]Corresponding author: zheminglu@zju.edu.cn

Received September 2015; accepted December 2015

Abstract. *This paper presents a scale-invariant approach to detect the interest points of images using complex network analysis. A network is built by considering each pixel of an image as a network node, and two nodes are connected if the distance of their corresponding pixels is not longer than a distance parameter $r$. On the basis that multiple values of $r$ can naturally allow a multi-scale analysis of an image, each image is represented as a set of complex networks with different values of $r$. After an integration step, interest points are spotted according to the normalized strength of each network node. Experiment results show the excellent performance of the proposed method under scale changes of images.*
**Keywords:** Interest point, Complex network, Scale-invariance, Image processing

1. **Introduction.** Feature detection is an essential step in computer vision and image processing. A feature is defined as an "interesting" part of an image, and features are used as a starting point for many computer vision algorithms. Interest points, namely the specific positions of some distinguishable points such as corners and edge points, usually served as the lowest-level features in an image. 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 and view-based object recognition. Interest point detectors have also been used as primitives for texture recognition, texture analysis and for constructing object models from multiple views of textured objects.

The best-known interest point detectors include the Moravec algorithm, the Harris and Stephens algorithm, the SUSAN detector and affine-adapted interest point operators [1-6]. In the Moravec algorithm [2,3], interest points are spotted according to the grey value differences between a window and windows shifted in several directions. Harris and Stephens [4] improve the approach of Moravec by using the auto-correlation matrix. The SUSAN detector [5] is based on the minimization of the local image region of each pixel and uses this region as the smoothing neighborhood for a noise reduction method.

One of the main difficulties of feature extraction is to obtain invariance to scale changes. Two scale-invariant detectors, are proposed in [6,7] by using the scale-space extrema in the LOG and DOG of images, respectively. Such methods have been demonstrated to be highly useful in practice, yet other approaches are still worth exploring.

Complex networks can be conceptualized as an interface between graph theory and statistical physics [8]. The study of complex network is an active area of scientific research inspired largely by the empirical study of real-world networks such as computer networks and social networks. One of the main reasons why complex networks have become so popular is their flexibility and generality in representing virtually any natural structure [9]. Thus, various studies include investigations on how to represent a problem as a complex network, followed by an analysis of its topological characteristics and feature extraction.

The use of tools and techniques of complex network analysis in computer vision is an appealing scientific topic that has been stated in recent years [10]. A general framework to integrate the area of computer vision research and complex networks is proposed in [11]. In [12,13], complex networks are used for modeling texture. In [14], an interest point detector using local centrality measures of corresponding complex networks has shown the value of complex network analysis in interest point detection.

In this paper we introduce a scale-invariant approach to detect the interest points of images using complex network analysis. An image is represented as a set of weighted complex networks that give information about image pixels and their neighbors. Each network has a different value of the distance parameter $r$, which allows a multi-scale representation of the image. Interest points are detected according to the node strength values after an integration step. This approach differs from the existing ones mainly in its complex network analysis and the scale-invariance of the points detected.

## 2. **Interest Point Detection Based on Complex Network Analysis.**

### 2.1. **General algorithm.**

For clarity, we first give the general algorithm before we explain the detailed process.

**Step 1 Network Generation.** Each image is associated with a set of networks with different values of the distance parameter $r$.

**Step 2 Node Strength Calculation.** For each image, calculate the strength of each node in all of its associated networks.

**Step 3 Integration of Information in Different Networks.** For each pixel in an image, compare the strength values of its corresponding node in all the associated networks, and only keep the maximum.

**Step 4 Detection.** Spot interest points according to the strength value of each node that is kept in the integration step.

### 2.2. **Network generation.**

From a mathematical point of view, a gray-level image $I$ of size $N \times N$ is a non-negative matrix $I = (I_{xy}) \in M_{N\times N}$ such that every entry $I_{xy}$ has an integer value (typically between 0 and 255 if we are dealing with 8-bit images), where $x$ and $y$ are the Cartesian coordinates of the pixel $I_{xy}$. One of the first complex network models for image was introduced in [11]. If $I$ is a gray-level image of $N \times N$ pixels, we can associate to it a weighted network $G = (X, E)$ with $|X| = N^2$ nodes so that each node corresponds to a pixel of $I$, and the weight of each link $(i, j) \in E$ is $w(i, j) = ||\vec{f_i} - \vec{f_j}||_2$, where $||\cdot||_2$ denotes the Euclidean norm and $\vec{f_i}$ is a feature vector that describes some local visual properties about the respective image pixels. Most of the classic interest point detectors are designed on the basis of the idea that interest points have high gradient values compared to their surrounding pixels. Thus, it is reasonable to set the weight of each edge to be the difference of pixel intensity between two nodes.

The main disadvantage of the idea introduced in [11] is the heavy computational cost largely because the network is always a complete weighted graph. To avoid this disadvantage, we introduce a distance parameter $r$: the nodes associated to two pixels are connected by an edge only when the Euclidean distance between their corresponding pixels is not longer than $r$. As a result, the complexity of generated networks is largely reduced, and the local nature of the interest points detected can be guaranteed.

2.3. **Combination of complex network analysis with multi-scale representation of images.** In fact, different values of $r$ naturally allow a multi-scale representation of the image. The rationale is that by using multiple values of $r$, information of local regions with different sizes can be obtained. In the case of scale changes, it is clear that a high value of $r$ is suitable for a high resolution image. Thus, we use a set of values of $r$ in our method: for each image, a set of networks are generated with $r = 2, 3, 4, \ldots, r_{\max}$.

**Remark 2.1.** *Note that we do not set r equal to 1. This is because when r equals 1, each node is only connected with the nodes corresponding to pixels in its 4-neighborhood in the image. In this case, only four directions parallel to the rows and columns of the image are used, causing inaccurate representation of the intensity differences between pixels.*

**Remark 2.2.** *The value of* $r_{\max}$ *is dependent on the resolution of the image. For example, if an image is shrunk by a factor of 2, then its* $r_{\max}$ *becomes* $r_{\max}/2$*. The ratio between the two* $r_{\max}$ *is exactly the scale factor between the images.*

**Remark 2.3.** *For a high resolution image, the use of low values of r is still necessary. Even images of high resolution have abrupt changes of pixel intensity, and thus we use all the integers from 2 to* $r_{\max}$ *as values of r.*

2.4. **Calculation of node strength.** The detection step is based on the detection of key nodes in networks. Many different methods have been proposed to spot key nodes in complex networks [15]. Most of these methods use the index of node degree, which is defined as the number of links incident upon a node. Since only weighted networks are generated in this work, we calculate the strength of each node in every network instead.

**Definition 2.1.** *Let* $G = (X, E)$ *be an undirected network. If* $i \in X$*, its node strength is given by* $s(i) = \sum_{(i,j)\in E} w(i, j)$*, where* $w(i, j)$ *is the weight of link* $(i, j)$*.*

Since the weight of each link is the difference of pixel intensity between two nodes, node strength is an appropriate measure of local property of images.

2.5. **Integration of information in different networks and final detection.** Recall that a set of networks with different $r$ are generated for each image, so we propose a method to integrate the information of all the networks together. We first normalize the strength of each node: for each network, we set $s(i) = (s(i) - s_{\min})/(s_{\max} - s_{\min})$, where $s_{\max}$ and $s_{\min}$ are the maximum and minimum of node strength of all the nodes in the network being considered, for every $i \in X$. Note that each node in a network corresponds to a pixel of the image, so networks with different $r$ have the same nodes regardless of the links. Thus, for each node, we can only keep its maximum strength value in all the networks.

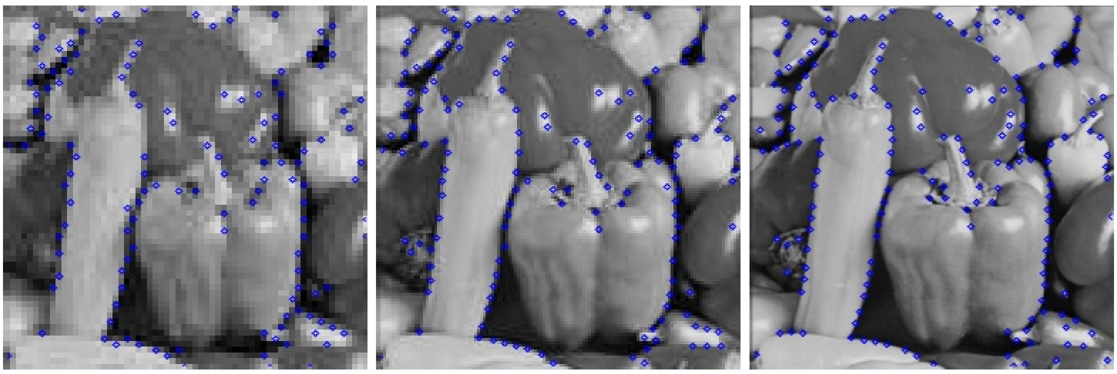
After the integration step, we can now spot the interest points of the image. We first pick out all the nodes whose strength is higher than a threshold value $T$. Since the strength of each node has been normalized, the value of $T$ is between 0 and 1. Then we rank these nodes according to their strength. We begin with the first one of these sorted nodes, and map them one by one back to the original image as interest points. To avoid the situation that too many redundant points are detected, a node is not mapped back if any pixel with distance less than $r_{\max}$ to its corresponding pixel in the image has been spotted as an interest point. At this point, the whole process of the proposed method is completed.

3. **Experimental Result and Analysis.** In this section, we compare the results of the Harris interest point detector [4] and the SIFT [7] with the results achieved by our method. We present the results of two known benchmark images Lena and Peppers. Images at lower resolutions are obtained by resizing the image at $256 \times 256$ using cubic

interpolation. For an image at a spatial resolution of $64 \times 64$, the value of $r_{\max}$ is 2. We multiple $r_{\max}$ by the scale factor for a image at different scales, and thus the value of $r_{\max}$ for a $256 \times 256$ image is 8. We use 0.4 as the value of threshold $T$ for all the images. The results of our method are shown in Figure 1 and Figure 2. From left to right, the resolution of the figure shown is $64 \times 64$, $128 \times 128$ and $256 \times 256$, and they are all resized to $256 \times 256$ for presentation.

(a) (b) (c)

Figure 1. Detection results of Lena at scale: (a) $64 \times 64$, (b) $128 \times 128$, (c) $256 \times 256$



(a) (b) (c)

Figure 2. Detection results of Peppers at scale: (a) $64 \times 64$, (b) $128 \times 128$, (c) $256 \times 256$

The results can show that the proposed method has very excellent performance under scale changes. For example, the end of nose and the inside corner of Lena's right eye are detected in all of the three images in Figure 1; each of the two white regions in the upper-middle part of Peppers has exact two interest points in all of the three images in Figure 2. Although these points may not be detected at the same position, they only have slight displacements in images at different scales.

Comparisons between our method with the Harris detector and the SIFT are also presented for scale changes. We use the repeatability criterion introduced in [3]. The set of point pairs $(x_1, x_i)$ which correspond with an $\varepsilon$-neighborhood is defined by:

$$R_i(\varepsilon) = \{(x_1, x_i) | dist(H_{1i}x_1, x_i) < \varepsilon\} \quad (1)$$

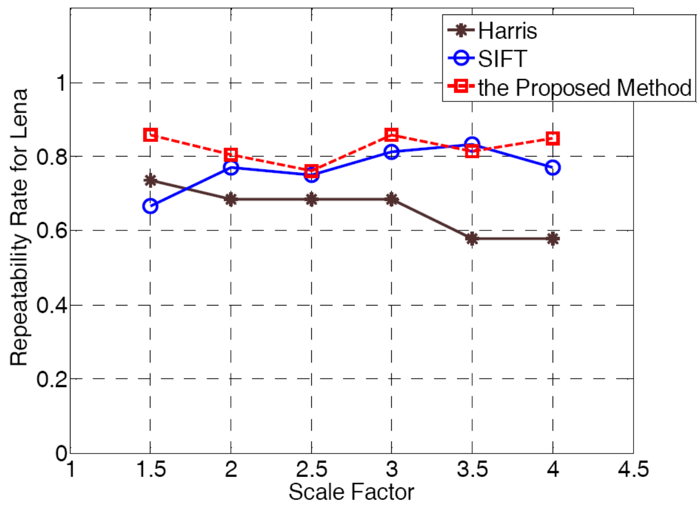
where $x_1$ is a point in image $I_1$ and $x_i$ is its corresponding point in image $I_i$, $I_i$ is the scaled version of $I_1$, and $H_{1i}$ represents the projection of $x_1$ to $x_i$. The repeatability rate

$r_i(\varepsilon)$ for image $I_i$ is defined by:

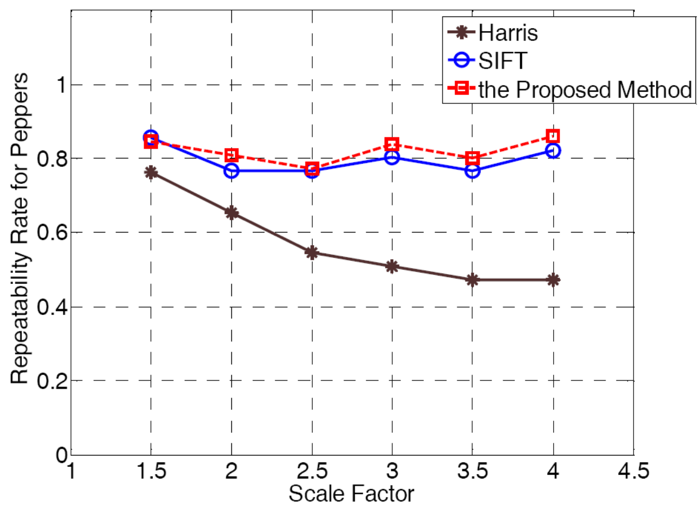
$$r_i(\varepsilon) = \frac{|R_i|}{\min(n_1, n_i)} \tag{2}$$

where $n_1$ and $n_i$ are the numbers of interest points detected in image $I_1$ and image $I_i$ respectively. Since more interest points are detected on the high resolution image, only the minimum number of interest points can be repeated. Thus, we use the image at the resolution of $64 \times 64$ as the reference image when we calculate the repeatability rate.

Figure 3 shows the repeatability rates of the three detectors. The scale factor is determined by the square-root of the ratio of the numbers of pixels of the image being considered and the reference image. We use 1.5 as the value of $\varepsilon$. For the Harris detector [4], the value of $k$ is 0.01, the size of the window for maximal suppression is $3 \times 3$, and we use the maximum of the response function multiplied by 0.01 as the threshold. For the SIFT, the parameters are the same as [7]. Evidently the Harris detector is very sensitive to scale changes, and its results are hardly usable. On the other hand, the repeatability rates of our method are mainly above 0.8, and are slightly higher than those of the SIFT.




(a)




(b)

FIGURE 3. Repeatability rates of the three detectors: (a) repeatability rates of Lena, (b) repeatability rates of Peppers

4. **Conclusion.** In this paper, a scale-invariant approach to detect the interest points of images using complex network analysis is proposed. Multiple values of the distance parameter $r$ are used to guarantee the scale invariance of the interest points. After an integration step, interest points are spotted according to the normalized strength of each node of the network. The experiment shows that the interest points detected by this method are well invariant to scale changes. In the future work, image matching methods can be studied based on the scale invariant results achieved by our method.

**Acknowledgment.** This work is partially supported by the following grants respectively: A Project of Shandong Province Higher Educational Science and Technology Program: J14LN25; Shandong Province Natural Science Foundation China: ZR2014FP003 and Project 61572269 supported by NSFC. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## REFERENCES

[1] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer Science & Business Media, 2010.
[2] H. P. Moravec, *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*, Stanford Univ. Ca. Dept. of Computer Science, no.STAN-CS-80-813, 1980.
[3] 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.
[4] C. Harris and M. Stephens, A combined corner and edge detector, *Alvey Vision Conference*, vol.15, pp.147-151, 1988.
[5] 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.
[6] K. Mikolajczyk and C. Schmid, Scale & affine invariant interest point detectors, *International Journal of Computer Vision*, vol.60, no.1, pp.63-86, 2004.
[7] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol.60, no.2, pp.91-110, 2004.
[8] M. E. J. Newman, The structure and function of complex networks, *Siam Review*, vol.45, no.2, pp.167-256, 2003.
[9] L. F. Costa, F. A. Rodrigues, G. Travieso et al., Characterization of complex networks: A survey of measurements, *Advances in Physics*, vol.56, no.1, pp.167-242, 2007.
[10] L. F. Costa, Complex networks: New concepts and tools for real-time imaging and vision, *arXiv Preprint cs/0606060*, 2006.
[11] L. F. Costa, Complex networks, simple vision, *arXiv Preprint Cond-mat/0403346*, 2004.
[12] T. Chalumeau, L. F. Costa et al., Texture discrimination using hierarchical complex networks, *Signal Processing for Image Enhancement and Multimedia Processing*, pp.95-102, 2008.
[13] A. R. Backes, D. Casanova and O. M. Bruno, Texture analysis and classification: A complex network-based approach, *Information Sciences*, vol.219, no.1, pp.168-180, 2013.
[14] R. Criado, M. Romance and Á. Sánchez, Interest point detection in images using complex network analysis, *Journal of Computational and Applied Mathematics*, vol.236, no.12, pp.2975-2980, 2012.
[15] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.