MULTISCALE KEYPOINT ANALYSIS WITH TRIANGULAR BIORTHOGONAL WAVELETS VIA REDUNDANT LIFTING

Kensuke Fujinoki

Department of Mathematical Sciences,
Tokai University
4-1-1 Kitakaname, Hiratsuka, Kanagawa, 259-1292 Japan
Phone: +81-463-58-1211
Email: fujinoki@tokai-u.jp

ABSTRACT

This paper presents an efficient approach for multiscale keypoint detection based on triangular biorthogonal wavelets. The detection scheme is simple and thus fast as only three isotropic directional components of an image obtained by multiscale decomposition with the triangular biorthogonal wavelets are used for keypoint localization at each scale. Redundant lifting is also considered and can be applied directly to calculate cumulative local energy distribution that is derived from the correction of the three directional components at each scale. This gives the efficient and accurate localization of keypoints including scale information. An experimental result shows that our method is better in the sense of the uniform distribution of keypoints compared with the conventional wavelet-based approach.

Index Terms— Keypoint, discrete wavelet transform, triangular lattice, lifting, redundant transform

1. INTRODUCTION

Local features of an image are often represented as edges, corners, keypoints, and so on. Among them, keypoints are widely used in the field of computer vision for the purpose of object recognition, content-based image retrieval, and image classification. While various techniques are used for keypoint detection, one notable method uses the scale-invariant feature transform (SIFT) [1] because it can describe multiscale keypoints that are robust to scale and rotation. This is different to the classical Harris corner detector [2], which detects corners as local features of an image. However, SIFT is highly computationally expensive because it computes the difference of Gaussian (DoG) at each scale to detect scale-invariant keypoints. It also needs to separate keypoints from edges.

This study focuses on wavelet-based approaches for multiscale keypoint analysis. Loupias et al. [3] proposed a fast salient points extraction method using the fast Mallat algorithm for multiscale signal decomposition with the discrete wavelet transform (DWT). However, the DWT is not shift-invariant and has a lack of directional selectivity. To account for this drawback, the dual-tree complex wavelet transform (DTCWT) has been successfully applied to multiscale keypoint detection [4]. It is directionally selective and approximately shift-invariant. However, this method uses complex bases, the constructions of the associated filters are highly involved, and it is computationally complex.

More recently, triangular wavelets that belong to a new class of two-dimensional wavelets defined on a regular triangular lattice have been proposed [5, 6]. They are easy to use because they are constructed using a straightforward generalization of one-dimensional wavelets. This means that they inherit several nice features of classical wavelets such as the systematic construction of filters and fast multiscale signal decomposition. They also maintain isotropy of an image, allowing isotropic image processing, as in the DTCWT. In addition, all of the associated filter coefficients are real and the filter construction is not computationally complex, which allows fast multiscale keypoint detection. Thus, we consider that these particular properties of the triangular wavelets are well suitable for keypoint analysis.

In this paper, we propose an efficient method for uniform keypoint detection based on triangular biorthogonal wavelets. We also apply a redundant lifting approach to construct wavelet filters that can be used directly in our keypoint detection scheme. The proposed methods provide an efficient and accurate localization of keypoints with scale information, which uniformly represent the main characteristics of an image unlike the conventional wavelet-based approach.

The remainder of this paper is organized as follows. In Section 2, we give an overview of the construction of triangular biorthogonal wavelets. Section 3 introduces the method for multiscale keypoint detection using triangular biorthogonal wavelets, implemented using redundant lifting. We show the results of the keypoint detection using some images with a comparison in Section 3. Finally, Section 4 gives conclusions.

2. TRIANGULAR BIORTHOGONAL WAVELETS USING LIFTING

Triangular wavelets are defined on a triangular lattice Λ, generated using a linear combination of two vectors $\boldsymbol{t}_1 = (1 \ 0)^T$ and $\boldsymbol{t}_2 = (-\frac{1}{2} \ \frac{\sqrt{3}}{2})^T$. The reciprocal lattice $\widetilde{\Lambda}$ that corresponds to the Fourier domain is similarly generated with vectors $\boldsymbol{\lambda}_1 = (0 \ \frac{2}{\sqrt{3}})^T$ and $\boldsymbol{\lambda}_2 = (1 \ \frac{1}{\sqrt{3}})^T$. For notational convenience, we also define $\boldsymbol{t}_0 = \boldsymbol{0}$, $\boldsymbol{t}_3 = -\boldsymbol{t}_1 - \boldsymbol{t}_2$, and $\boldsymbol{\lambda}_3 = \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2$.

The polyphase representation of a two-dimensional discrete signal, $c_j[\boldsymbol{t}], \boldsymbol{t} \in \Lambda$, defined on the lattice that has a resolution level $j \in \mathbb{Z}$ and $j \geq 0$, is written by

$$\hat{c}_j(\boldsymbol{\omega}) = \sum_{m=0}^{3} e^{-i\boldsymbol{\omega} \cdot \boldsymbol{t}_m} \, \hat{c}_{m,j}(2\boldsymbol{\omega}),$$

where

$$\hat{c}_{m,j}(\boldsymbol{\omega}) = \sum_{\boldsymbol{t} \in \Lambda} c_j[2\boldsymbol{t} + \boldsymbol{t}_m] e^{-i\boldsymbol{\omega} \cdot \boldsymbol{t}}, \qquad m = 0, 1, 2, 3.$$

This means that we have four polyphase components $\hat{c}_{m,j}(\boldsymbol{\omega})$, implying that we have one even component, $c_j[2\boldsymbol{t}]$, and three odd components, $c_j[2\boldsymbol{t} + \boldsymbol{t}_m], m = 1, 2, 3$. Note that in one dimension there only exist even $c_j[2k], k \in \mathbb{Z}$ and odd $c_j[2k+1]$ components.

Following a straightforward generalization of the DWT, a signal $c_j[\boldsymbol{t}]$ is decomposed into a coarse component $c_{j+1}[\boldsymbol{t}]$ and three detail components $d_{k,j+1}[\boldsymbol{t}], k = 1, 2, 3$, of half resolution. This can be written using the lifting form [7] generalized to the two dimensional lattice. Three detail components are obtained using three predictors $p_k, k = 1, 2, 3$, that is,

$$d_{k,j+1}[\boldsymbol{t}] = c_j[2\boldsymbol{t} + \boldsymbol{t}_k] - p_k(c_j[2\boldsymbol{t}]), \qquad k = 1, 2, 3. \quad (1)$$

A coarse component is derived using the results of the prediction with updater $u_k, k = 1, 2, 3$,

$$c_{j+1}[\boldsymbol{t}] = c_j[2\boldsymbol{t}] + \sum_{k=1}^{3} u_k(d_{k,j+1}[\boldsymbol{t}]). \quad (2)$$

This preserves the average of a two-dimensional signal. Finally, the normalization steps are applied to normalize the L^1 norm. The triangular DWT that repeats the predict (1) and update (2) steps up to a resolution level $L > j$ produces the multiscale coefficients

$$c_j[\boldsymbol{t}] \rightarrow \{d_{k,j+1}[\boldsymbol{t}], d_{k,j+2}[\boldsymbol{t}], \dots, d_{k,L}[\boldsymbol{t}], c_L[\boldsymbol{t}]\}, \\ k = 1, 2, 3. \quad (3)$$

Different choices of p and u give an arbitrary set of biorthogonal wavelet filters. For example, the simplest choice

$$\hat{p}_k(\boldsymbol{\omega}) = 1, \qquad \hat{u}_k(\boldsymbol{\omega}) = 1/4, \qquad k = 1, 2, 3, \quad (4)$$
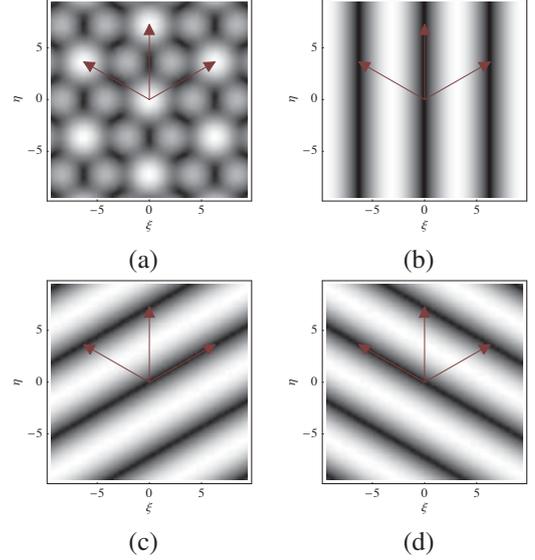

Fig. 1. Frequency responses of triangular biorthogonal Haar wavelet filters (h, g_k). (a): $\hat{h}$. (b):$\hat{g}_1$. (c): $\hat{g}_2$. (d): $\hat{g}_3$. The arrows indicate 2π times vectors $\boldsymbol{\lambda}_k, k = 1, 2, 3$.

gives a system of triangular biorthogonal Haar filters that consists of a low-pass (LP) filter $h[\boldsymbol{t}]$ and three high-pass (HP) filters $g_k[\boldsymbol{t}]$, and their duals $\tilde{h}[\boldsymbol{t}]$ and $\tilde{g}_k[\boldsymbol{t}]$. Note that we have three independent primal HP filters $\hat{g}_k$ and dual HP filters $\hat{\tilde{g}}_k$, because we have three predictors p_k. As in (4), if we set the same predictors for the three prediction steps in (1), the resulting filters become isotropic. This means that the HP filters $\hat{g}_k$ and $\hat{\tilde{g}}_k$ have directional responses to the $\boldsymbol{t}_k$ directions on the reciprocal lattice. Thus, each HP filter is symmetrically defined by rotating it by $\pm 2\pi/3$ on the lattice (see Fig. 1).

3. MULTISCALE KEYPOINT DETECTION

The triangular DWT (3) produces three detail components $(d_{k,j}[\boldsymbol{t}], k = 1, 2, 3)$ which contain directional information for each scale. While the DTCWT decomposes a signal into six orientations of subbands in the frequency domain, each $d_{k,j}[\boldsymbol{t}]$ reveals the edge structure of an image in the directions towards $0°, 120°$, and $240°$, because the three HP filters are designed to be isotropic. In fact, these three directions are very efficient when analyzing the directional components of a signal, because they form a hexagon that is the most symmetric polygon that can fill up a two-dimensional plane $\mathbb{R}^2$.

To extract the keypoints from the detail components, the local energy distribution is defined by

$$\tilde{d}_j[\boldsymbol{t}] = \alpha^j \prod_{k=1}^{3} |d_{k,j}[\boldsymbol{t}]|^2, \quad (5)$$

which represents the unique points of images where the three edges of the detail components intersect at each scale. The

parameter α adjusts a magnitude scale of $\tilde{d}_j[\boldsymbol{t}]$, which distinguishes the features of an image from other components such as noise. Note that $d_{j,k}[\boldsymbol{t}]$ is assumed to be smoothed by the Gaussian filter, which means that we can more easily detect its peaks and reduce the influence of noise.

3.1. Redundant lifting

The local energy distribution $\tilde{d}_j[\boldsymbol{t}]$ at each scale level j produced by the standard lifting has the different location that depends on j. Therefore, we introduce redundant lifting to obtain the same index ($\boldsymbol{t}$) for each scale of the local energy distribution $\tilde{d}_j[\boldsymbol{t}]$, which allows to directly combine the $\tilde{d}_j[\boldsymbol{t}]$ of each scale. The previously described lifting uses the polyphase decomposition, which classifies a signal $c_j[\boldsymbol{t}]$ into one even $c_j[2\boldsymbol{t}]$ and three odd components $c_j[2\boldsymbol{t} + \boldsymbol{t}_k]$. As a result, each polyphase component has half the resolution of the original signal. The redundant lifting does not use the polyphase decomposition or decimation.

In the case of the Haar filter, which uses constant predictions $\hat{p}_k(\boldsymbol{\omega}) = 1$ and updates $\hat{u}_k(\boldsymbol{\omega}) = 1/4$, the steps can be rewritten as

$$d_{k,j+1}[\boldsymbol{t}] = c_j[\boldsymbol{t} + 2^j\boldsymbol{t}_k] - c_j[\boldsymbol{t}], \qquad k = 1, 2, 3, \quad (6)$$

and

$$c_{j+1}[\boldsymbol{t}] = c_j[\boldsymbol{t}] + \sum_{k=1}^{3} \frac{d_{k,j+1}[\boldsymbol{t}]}{4}, \quad (7)$$

where j starts from 0. A similar setting to the one for $c_j[k + 2^j]$ can also be seen in the à trous algorithm [9], which realizes the undecimated Mallat transform. Therefore, the redundant lifting described here is different from that of [8], which applies the prediction and update steps twice to implement the redundant transform. Unlike the standard prediction, the redundant lifting maintains the resolution of the decomposed signals.

Here we describe how redundant lifting decomposes a signal. Before applying the decomposition, we use the simple half-shift pixel method [10] to generate a triangular lattice from a square sample of image data. For each odd line on a square lattice, we first find the midpoint between two adjacent pixels using linear interpolation. Then, we discard the left and right, keeping only the mid values. As a result, we obtain a hexagonally sampled triangular lattice. We use redundant lifting to decompose the original image into a coarse approximation $c_{j+1}[\boldsymbol{t}]$, and three oriented detail components, $d_{k,j+1}[\boldsymbol{t}], k = 1, 2, 3$, at each resolution level. Note that they have the same resolution density.

3.2. Cumulative local energy distribution

The locations of the local energy distribution at each scale j are maintained, because the $d_{k,j+1}[\boldsymbol{t}]$ obtained by the redundant lifting decomposition has exactly the same index space.
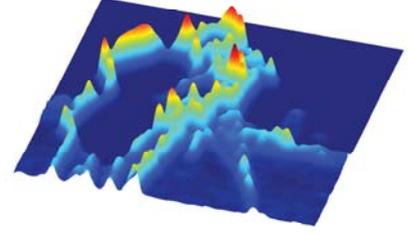
(a)

(b)

Fig. 2. (a) Cumulative local energy distribution. (b) 40 keypoints.

To add scale information to the local energy distribution, we define a cumulative local energy distribution as

$$P[\boldsymbol{t}] = \sum_{j=1}^{L} \tilde{d}_j[\boldsymbol{t}]. \quad (8)$$

The multiscale keypoints are obtained by simply detecting the local maxima in the cumulative local energy distribution $P[\boldsymbol{t}]$, and their locations are the same as the indices $\boldsymbol{t}$ because of the redundant lifting.

Fig. 2(a) shows the cumulative local energy distribution $P[\boldsymbol{t}]$ of the cameraman image, and Fig. 2(b) shows the result of the keypoint detection where 40 keypoints were detected using the cumulative local energy distribution. The distribution is smooth because of the Gaussian filtering, and thus the keypoints of neighboring local features do not overlap. In Fig. 2(b), each circle on the image represents the scale of the keypoints, which corresponds to the value of the local maxima of $P[\boldsymbol{t}]$. As we intended, all the keypoints represent distinguishing local features of the image such as corners and junctions. This implies that three symmetric edge features extracted using triangular biorthogonal wavelets accurately represent the local characteristics of an image, and the scale information of the keypoints provide its global property.

(a)   (b)

(c)   (d)
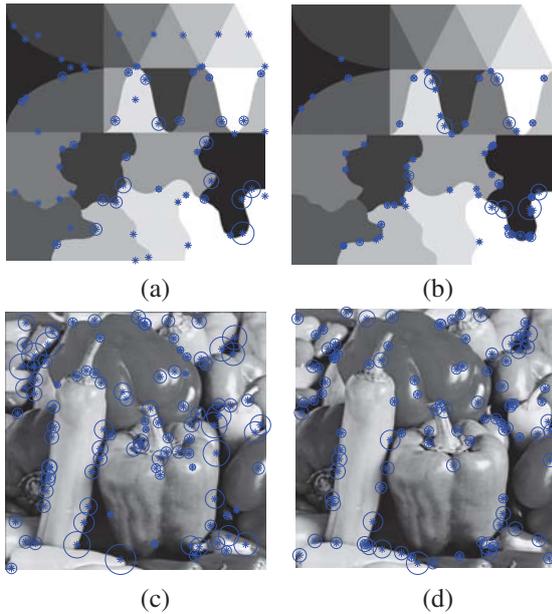
Fig. 3. Results of the keypoint detection with $P = 64$ (top) and $P = 100$ (bottom). Left column shows the case of the triangular wavelet; right column shows the conventional wavelet. Each entropy E is given by (a) 95.6%, (b) 82.0%, (c) 91.6% and (d) 88.4%, respectively.

3.3. Uniform detection of keypoints

Here we evaluate the performance of the proposed method when compared with the conventional discrete wavelet-based approach [3]. Because our triangular wavelet filters maintain the isotropy of images, we expect that the keypoints will be uniformly detected if an image is isotropic. To measure the uniformity of the distribution of keypoints, we divided the image into N blocks. We then calculated the entropy in each k-th block using

$$E = \frac{-\sum_{k=1}^{N} p[k] \log(p[k])}{\log(N)}. \tag{9}$$

This represents the occurrence rate of p keypoints existing in each block k, when an image is divided into N equal blocks [3]. A larger E indicates that the keypoints are more uniformly distributed.

Fig. 3 shows the results of the keypoint detection for two images containing nearly isotropic edges, and their entropies when $P = N = 64$ and $P = N = 100$. These results demonstrate the advantage of our method when compared with the conventional technique. When using triangular wavelets, the keypoints are more uniformly distributed, and consequently the entropy is higher. This is because the triangular wavelets uniformly decompose an image and preserve its isotropy.

4. CONCLUDING REMARKS

We proposed a method for multiscale keypoint detection based on triangular biorthogonal wavelets. Our method uses the cumulative local energy distribution of three isotropic detail components obtained from redundant lifting on the triangular lattice to detect the distinctive local features of an image and their locations. The extracted keypoints include scale information that represents the local and global characteristics of an image.

The proposed method is efficient and accurate, because it only uses three symmetric edge components for detecting keypoints. Additionally, it produces a more uniform distribution of keypoints when compared with the conventional wavelet-based approach due to the isotropic signal decomposition. The keypoint descriptor should be developed in future work, and further comparisons of the performance when compared with recently developed methods are necessary.

REFERENCES

[1] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, Vol. 60, No. 2, 91–110, 2004.

[2] C. Harris and M. Stephens, A combined corner and edge detector, Alvey Vision Conference, 147–151, 1988.

[3] E. Loupias, N. Sebe, S. Bres, and J-M. Jolion, Wavelet-based salient points for image retrieval, IEEE International Conference on Image Processing, 2000.

[4] J. Fauqueur, N. Kingsbury, and R. Anderson, Multi-scale Keypoint Detection Using the Dual-Tree Complex Wavelet Transform, IEEE International Conference on Image Processing, pp. 1625–1628, 2006.

[5] S. Sakakibara and O. V. Vasilyev, Construction of triangular biorthogonal wavelet filters for isotropic image processing, EUSIPCO2006, September, 2006.

[6] K. Fujinoki and O. V. Vasilyev, Triangular wavelets: an isotropic image representation with hexagonal symmetry, EURASIP Journal on Image and Video Processing, Vol. 2009, No. 248581, 16 pages, 2009.

[7] W. Sweldens, The lifting scheme: a custom-design construction of biorthogonal wavelets, J. Appl. Comput. Harmon. Anal., Vol. 3, No. 2, pp. 186–200, 1996.

[8] R. Claypoole, R. Baraniuk and R. Nowak: Adaptive wavelet transform via lifting, ECE TR 9304, 1999.

[9] M. J. Shensa, Discrete wavelet transforms: Wedding the à trous and Mallat algorithms, IEEE Transactions on Signal Processing, 40, 2464-2482 (1992)

[10] A. Laine and S. Schuler, Hexagonal wavelet processing of digital mammography, Proceedings of SPIE, Vol. 1898, Image Processing(193)/57.