Advances in Computer Science Research, (ACSR), volume 52
2016 International Conference on Computer Engineering and Information Systems (CEIS-16)

Complex NSST in Text Localization

Jiu-Wen Zhang, Lu Zhang, Chun-Hui Liu
School of Information Science and Engineering, Lanzhou University,
Country222 Tianshui Road, Lanzhou, P.R. China
E-mail: zhangjw@lzu.edu.cn, Zhanglu2014@lzu.edu.cn, liuchh14@lzu.edu.cn

Min Dong
School of Information Engineering, Zhengzhou University,
No.100 Science Avenue, Zhengzhou, P.R. China
E-mail: iemdong@zzu.edu.cn

***Abstract*-Text localization is a crucial course for image processing. In this paper, an algorithm for text localization based on Complex Nonsubsampled Shearlet transform (NSST) is proposed. Complex NSST is implemented by a projection followed NSST. The Projection operation converts the input image into its analytic one, and NSST composes the real component and imaginary component into multiscale and multidirectional subbands. The obtained relative phase and magnitudes are combined to get the new subbands. Then Dynamic thresholding, Morphological operations, logical AND operation, Voting-decision and Vertical and Horizontal Projections are used on the new subbands to achieve the final text regions. Compared with common wavelet transform, the complex wavelet transform has the properties of shift invariance, limited redundancy and low aliasing. At the same time, the added phase information delivers more accurate representation for image. Experiments show that the algorithm can accurately locates the text regions in an image.**

Keywords-text localization; complex wavelet transform; NSST; projection; relative phase

I. INTRODUCTION

Text contents in images contain important information that can represent the main details or mask the category of images. Text regions in images may be affected by brightness, shadow, typeface, size, pattern and complex background, which make the localization be a challenging thing. A large number of text localization algorithms have been proposed [1]. Yin et al. presented an approach by using MSER, geometric features and adaboost [2] which completed localization of text contents in images. Neumann et al. worked on text detection with Oriented Stroke Detection and an unconstrained end to end method [3]. Neha et al.[4] used Soble edge detector on three detail components which were obtained from Discrete Wavelet Transform, the edge map used for text localization was formed by resultant edges so obtained. Zhang et al.[5] applied Discrete Shearlet Transform on an image to obtain the directional subbands in different orientations and scales, the subbands with texture details were used to recognize the edges of text regions by filtering out the backgrounds.

In text localization, directional multiresolution is a useful and widly used tool. But traditional real wavelet transform, such as DWT, contourlet and shearlet, suffers from three shortcomings: 1) the coefficients derived from real wavelet transform tend to oscillate positive and negative around singularities which complicates wavelet-based processing, 2) shift variance also complicates wavelet-domain processing, and 3) wavelet coefficients obtained from iterated discrete-time downsampling operations will result in substantial aliasing. Through some researches, we find that complex wavelet transform [6] can overcome these shortcomings. Furthermore, we can achieve phase information which reveals positions of bump margin from complex wavelet subbands. For text locating, the phase information can substantially affect the results.

In this paper the complex wavelet transform is accomplished by combining a projection operating [7] and NSST [8]. For a real signal, Projection is adopted to create its analytic signal which can get precise Hilbert pair. The NSST is used to decompose an image into several directions and scales. The obtained subbands are consisting of some high frequency subbands and a low frequency subband which mainly contain text information and background information, respectively. We obtain amplitude and phase information from the subbands of the complex NSST and combining them into new subbands. Then text regions are screened and backgrounds information are suppressed with the help of Dynamic thresholding, Morphological operations, logical AND operation, Voting-decision and Vertical and Horizontal Projections. Experiments show that, the algorithm can improve the accuracy of locating text regions.

II. PROJECTION AND NSST

A. NSST

Shearlet is developed based on the composite wavelets which is the combination of Geometric Analysis and multiresolution analysis through Classical affine systems [9]. The formula of shearlet is derived as followed:

$$\left\{\psi_{a,s,t}(x)=a^{-3/4}\psi\left(A_a^{-1}B_s^{-1}(x-t)\right),a\in R^+,s\in\Re,t\in Z^2\right\} \quad (1)$$

where $A_a=\begin{bmatrix}\alpha & 0\\ 0 & \sqrt{\alpha}\end{bmatrix}$, $B_s=\begin{bmatrix}1 & -s\\ 0 & 1\end{bmatrix}$ and A_a is an anisotropic dilation matrix which realizes the multiscale change with the change of a, B_s is a shear matrix which keeps the area unchanged with the change of s, different values of t let shearlet be translation invariant.

The tiling of the frequency by shearlet with different values of *a* and *s* is illustrated in Fig.1.

The model of discrete shearlet transform can be produced by choosing a proper discrete set of *a* and *s* . NSST is constitutive of scale decomposition and direction localization, the decomposition and localization are accomplished by non-subsampled pyramid (NSP) and shearing filters (SF), respectively. NSP decompose the input image into a high-frequency sub-image and a low-frequency sub-image whose sizes are same as the input image. Let *S* be the decomposition levels, we will get $S+1$ sub-images which consist of *S* high-frequency components and one low-frequency component. SF will finish the multidirectional decomposition on each high-frequency component. Fig.2 shows the mtltiscale and multidirectional decomposition with $S=3$.

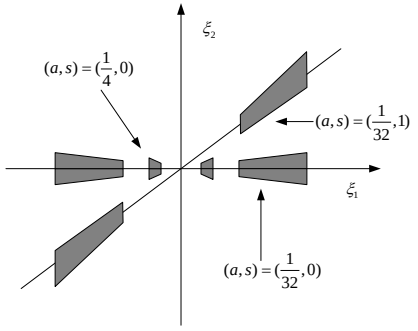


Figure 1. The structure of frequency tiling by Shearlet.

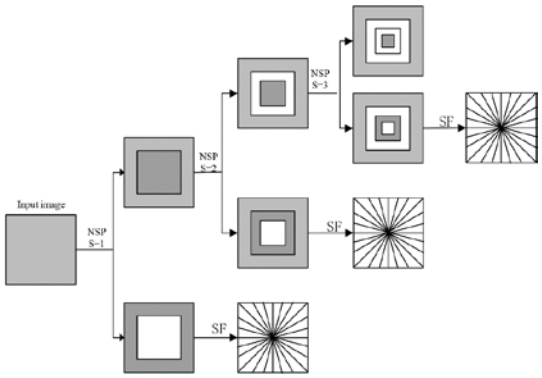
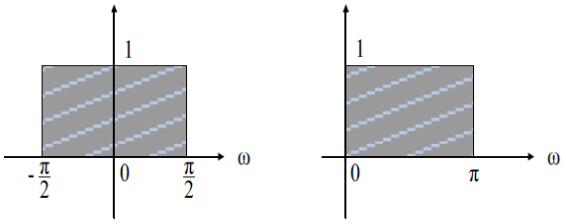


Figure 2. The decompositions of NSST.

B. Projection Filter

An analytic signal $f(x)$ is computed as $f(x)=f_R(x)+if_I(x)$, where $f_I(x)$ has the Hilbert transform relationship with $f_R(x)$. Thus $f_R(x)$ and $f_I(x)$ can be considered as the real part and the imaginary part of $f(x)$, respectively. In [7], a complex wavelet transform which is based on projection is proposed by Fernandes et al. The original signal is projected onto a signal space to obtain its real and imaginary parts which form a Hilbert pair and then an implementation of projection filters is proposed. The projection filter is achieved by shifting the low-pass analysis and synthesis filter of maximally flat wavelet filters by $\pi/2$ at phase as in [10].

Fig.3 shows the amplitude response for an ideal low-pass filter and ideal projection filter. For real-valued two-dimensional images, the negative half-plane of the Fourier transform carries redundant information. By filtering the rows or columns of the image with the projection filter, the redundant information is eliminated.



(a) (b)

Figure 3. Projection filter (a) ideal half-band low-pass filter (b) ideal projection filter.

C. Complex NSST

Motivated by [7], we construct complex NSST based on projection and NSST. Complex NSST first projects a real-valued signal into an analytic signal, and then NSST is applied on the obtained real and imaginary components, separately. Through complex NSST we can obtain a set of real parts subbands and a set of imaginary parts subbands which have the same size with the original image, and each set of subbands contain one low frequency subband and *n* high frequency subbands. The magnitude information is obtained directly from these subbands. We set the real and imaginary subband coefficient at a certain scale and direction as $F_R(x)$ and $F_I(x)$, respectively, and a new set of amplitude coefficients is obtained as followed:

$$F(x)=\sqrt{F^2{}_R(x)+F^2{}_I(x)} \tag{2}$$

The relative phase (RP) [11] matrix for coefficient $c_{sk}(i,j)$ can be obtained from each complex subband as followed:

$$RP_{sk}(i,j)=\begin{cases}\angle c_{sk}(i,j)-\angle c_{sk}(i,j+1) \ \ if\ 1\le k\le \dfrac{K}{2},\\ \angle c_{sk}(i,j)-\angle c_{sk}(i+1,j) \ \ if\ \dfrac{K}{2}<k\le K.\end{cases} \tag{3}$$

where $\angle$, (i,j), *s* and *k* represents the phase, position, scale and orientation of coefficient, respectively, and *K* is the total number of orientations. The scheme of getting magnitudes and RP from complex NSST is shown in Fig.4.

As the magnitudes and phases of complex wavelets coefficients indicate the strengths and locations of variations existing in original image, respectively, the combination of magnitudes and phases can be used to locate the text region exactly. The combination of magnitudes and relative phase is completed through $F(x)*RP$, which is used as subband

coefficient in the follow-up processing. Fig.5 shows the original input image, the analytic components, the multiscale and multidirectional subbands and one of the combined new subbands.

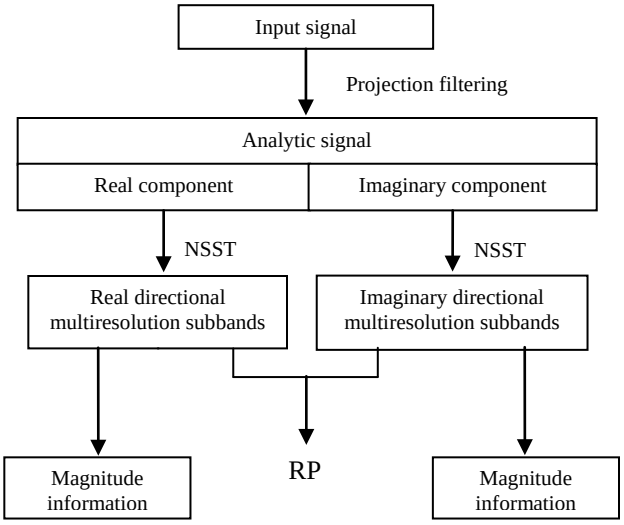


Figure 4. The scheme of getting magnitude and relative phase from complex NSST.

III. PROPOSED METHOD FOR TEXT LOCALIZATION

Before complex NSST decomposing, the input images must be converted into gray images. The main steps of text localization are demonstrated as followed:

1) Multiscale and multidirectional decomposition utilizing complex NSST;

2) Obtain the magnitudes and relative phase information from subbands of complex NSST and get the new combined subbands;

3) An appropriate threshold value to remove the non-text edges in high frequency subbands;

4) Connect the useful edges together in each sub-image by morphological operations;

5) Logical AND operation to the subbands in all directions for each scale;

6) Voting-decision among different scales to get candidate text pixels image;

7) Vertical and Horizontal Projections of candidate text pixels image to get text regions.

A. Dynamic Thresholding

Complex NSST decomposes an image into some high frequency subbands and a low frequency subband which mainly contain text information and background information, respectively. The high frequency subbands mainly contain text information, but there is also little non-text information in these subbands. And the removing of non-text pixels is realized by binarization operation, namely let the value of text pixels and non-text pixels be 1 and 0, respectively. Hence the intensity values of text region and non-text region are different, text pixels can be distinguished from non-text pixels by an applicable threshold value.

In our method, the applicable threshold value called T is developed by utilizing the dynamic thresholding analysis in [12]. T is built based on each pixel and its neighborhood pixels, the formulae is given as followed:

$$T = \frac{\sum(|es(i,j)| \times s(i,j))}{\sum s(i,j)} \tag{4}$$

$$s(i,j) = \max\begin{pmatrix} |es(i-1,j) - es(i+1,j)|, \\ |es(i,j-1) - es(i,j+1)| \end{pmatrix} \tag{5}$$

where $es(i,j)$ is the pixel value of a detail subband in position (i,j), and let the size of subband be $[M,N]$, $i = 1,2,\cdots M$, $j = 1,2,\cdots N$.

For pixels in the edges of subbands, that is for $i = 1$, $i = M$, $j = 1$, $j = N$, we set $s(i,j) = 0$. Then results of binarization operation eb is calculated as followed:

$$eb(i,j) = \begin{cases} 1 & es(i,j) \geq T \\ 0 & es(i,j) < T \end{cases} \tag{6}$$

where eb is the desired subband which is obtained from dynamic thresholding analysis. One of the multiscale and multidirectional subbands after Dynamic thresholding is shown as Fig.6(a).

B. Morphological Operation, Logical and Operator and Voting-decision

The methods of Morphological Operation, Logical AND Operator and Voting-decision we used come from the researches in [5]. Fig.6(b) shows one of the multiscale and multidirectional subbands after Morphological Operation. One of the multiscale subbands after Logical AND Operator is shown as Fig.6(c). Candidate text pixels image after Voting-decision is shown as Fig.6(d).

C. Vertical and Horizontal Projections

Vertical And Horizontal Projections convert the candidate text pixels image into one-dimensional representations. They are defined as [13]:

$$P_{hor}(i_0) = \sum_{j=0}^{N} es(i_0, j),\ for 0 \leq i_0 \leq M \tag{7}$$

$$P_{ver}(j_0) = \sum_{i=0}^{M} es(i, j_0),\ for 0 \leq j_0 \leq N \tag{8}$$

where $[M, N]$ is the size of the candidate image, P_{hor} and P_{hor} is the horizontal projection and vertical projection, respectively. Then two thresholds are used to screen text regions, they are defined as:

$$T_h = \frac{mean(P_{hor}) + \min(P_{hor})}{2} \tag{9}$$

$$T_v = \frac{mean(P_{ver}) + \min(P_{ver})}{2} \tag{10}$$

If $P_{hor}(i_0)$ is greater than T_h, the i_0 row can be considered as a candidate text region; otherwise, this row is suppress. If $P_{ver}(j_0)$ is greater than T_v, the j_0 column can be considered as a candidate text region. Then the horizontal and vertical image are combined as the final text regions image, which is shown as Fig.6(e).

IV. EXPERIMENTAL RESULT

All the experiments are implemented in MATLAB R2012a with an Intel core 3 CPU 2.53GHz machine. The MSRA-TD500 dataset has been used for text localization. The experiments are compared with NSST. We decompose each image in 3 scales and 6 directions at each scale, the filter's size vector is [6 6 6].

The results are shown as Fig.7. It can be seen that the proposed method performs well and can locate text regions more correctly than NSST. This conforms the properties of complex wavelet transform. The accurate localizations indicate that relative phase and the combination with magnitude are useful feature.

V. CONCLUSION

In this paper, an algorithm for text localization based on complex NSST is proposed. Complex NSST first decomposes the input image into multiscale and multidirectional subbands, relative phase and magnitudes of these subbands are combined to obtain the new subbands. Then Dynamic thresholding, Morphological operations, logical AND operation, Voting-decision and Vertical and horizontal projections are used to remove the non-text regions. This algorithm uses Projection to accomplish complex wavelet transform. The combination of relative phase and magnitudes has an accurate representation of image for text localization. Experimental results show that, the algorithm can improve the accuracy of text localization and suppress more backgrounds information.

REFERENCES

[1] Ye Q. and Doermann D., "Text Detection and Recognition in Imagery: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, vol. 37(7), pp. 1480-1500.

[2] Yin, X. C., Hao, H. W., and Iqbal, K., "Effective text localization in natural scene images with MSER, geometry-based grouping and AdaBoost, " In Pattern Recognition (ICPR 12), Nov. 2012 , pp. 725-728.

[3] Neumann L and Matas J, "Scene text localization and recognition with oriented stroke detection," Proceedings of the IEEE International Conference on Computer Vision. 2013, pp. 97-104.

[4] Gupta N and Banga V K, "Localization of Text in Complex Images Using Haar Wavelet Transform," International Journal of Innovative Technology and Exploring Engineering (IJITEE 12), 2012, pp. 2278-3075.

[5] Zhang J and Chong Y, "Text localization based on the Discrete Shearlet Transform," Software Engineering and Service Science (ICSESS 13), 2013, pp. 262-266.

[6] Selesnick I W, Baraniuk R G and Kingsbury N C, "The dual-tree complex wavelet transform," IEEE signal processing magazine, 2005, vol. 22(6), pp. 123-151.

[7] Fernandes, F., van Spaendonck, R., Coates, M. J., and Burrus, C. S. , "Directional complex-wavelet processing," International Symposium on Optical Science and Technology. International Society for Optics and Photonics, 2000, pp. 536-546.

[8] Kong W, "Technique for gray-scale visual light and infrared image fusion based on non-subsampled shearlet transform," Infrared Physics and Technology, 2014, vol. 63(11), pp. 110-118.

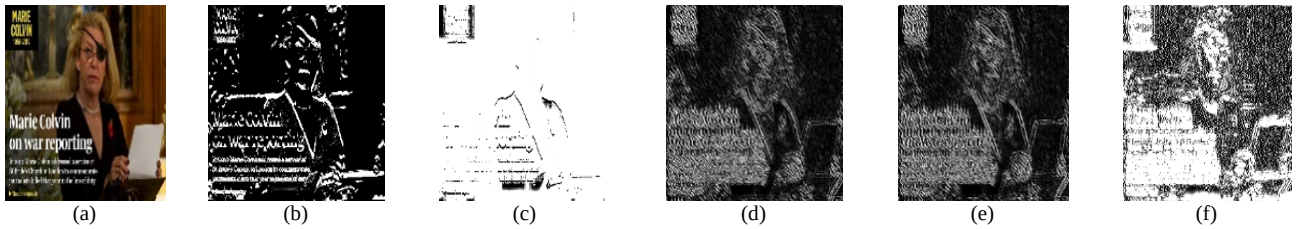
[9] Guo K and Labate D, "Optimally sparse multidimensional representation using shearlets," SIAM journal on mathematical analysis, 2007, vol. 39(1), pp. 298-318.

[10] Selesnick I W, "Low-pass filters realizable as all-pass sums: design via a new flat delay filter," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 1999, vol. 46(1), pp. 40-50.

[11] Vo, An, and Soontorn Oraintara, "A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation," Signal Processing: Image Communication, 2010, vol. 25(1), pp. 28-46.

[12] Lee S U, Chung S Y and Park R H, "A comparative performance study of several global thresholding techniques for segmentation," Computer Vision, Graphics, and Image Processing, 1990, vol. 52(2), pp. 171-190.

[13] Burger, W., Burge, M. J., Burge, M. J., and Burge, M. J., "Principles of Digital Image Processing," London: Springer, 2009, pp. 221.

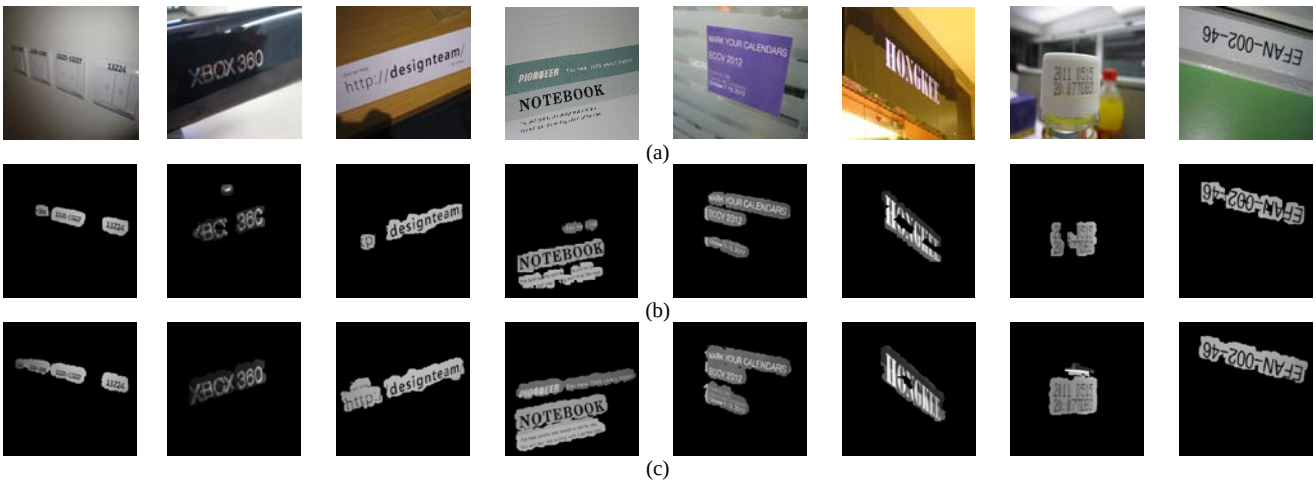


(a) (b) (c) (d) (e) (f)

Figure 5. A sample of image with complex NSST (a) input image (b) real component (c) imaginary component (d) one real subband (e) one imaginary subband (f) one new combined subband.

(a) (b) (c) (d) (e) (f)

Figure 6. A sample of result after each step (a) Dynamic thresholding (b) Morphological Operation (c) Logical AND Operator (d) Voting-decision (e) Vertical And Horizontal Projections (f) final result.



(a)

(b)

(c)

Figure 7. Experiment results (a) input images (b) NSST (c) complex NSST.