Extraction Algorithm of English Text Information From Color Images Based on Radial Wavelet Transform

YAQIN WANG
Department of Basic Courses, Ordos Vocational College, Ordos 017000, China
wangyaqinmail@163.com

ABSTRACT The Chinese and English text information of color image reflects some important contents of color image to a certain extent, and these texts are extracted automatically, which has important application value in the fields of information retrieval, digital library, web page retrieval and intelligent transportation. In this paper, various signal extraction techniques based on radial wavelet transform modulus Maxima are analyzed, and it is found that these techniques have poor ability to extract weak signals, and the higher requirements for directionality lead to pseudo-boundary phenomena in two-dimensional image extraction results. Based on the correlation denoising method of radial wavelet coefficients, the radial wavelet entropy is introduced into the field of signal extraction, and the complex Morlet radial wavelet is selected as the basis function. A complex Morlet radial wavelet entropy extraction algorithm suitable for extracting the English text characteristics of weak signal color images is proposed. In addition, a method of scene text recognition is proposed. Based on D-SIFT local features and spatio-temporal histogram, this method vectorizes text samples in the framework of Bo Ws model. Because the spatio-temporal histogram can flexibly model the structural information of the text, this method can effectively describe the scene text. On this basis, the selective ensemble learning method is used to improve the performance of the text extractor. In order to expand the application range of the algorithm and improve the running efficiency of the integrated extractor, a model compression method based on color image English text samples and integrated extractor is proposed. The integrated extractor which takes up more space and runs slowly is compressed into an equivalent and more efficient one. In order to reduce the number of pseudo-samples needed in the process of model compression, a method of training local extractor based on color image English text samples is proposed, which greatly reduces the number of pseudo-samples needed and improves the efficiency of the model compression method.

INDEX TERMS Color image, radial wavelet transform, wavelet basis function, English text information extraction, model compression.

I. INTRODUCTION
With the rapid development of the Internet and the massive accumulation of network resources, multimedia database has been greatly enriched, so content-based image analysis technology has attracted more and more attention [1]. At the same time, a large number of images and videos often contain certain English text information, and the English text information in these images reflects part of the important content of the image or video to a certain extent. If it can be accurately extracted and recognized, it can not only be used in content-based image retrieval systems, but also assist in the design of automatic navigation vehicles and artificial vision systems, visual impairment aids, text translation systems, mobile phone assistance systems and spam filtering systems [2]–[4]. For example, an image English text recognition module, a translation module and an Internet retrieval module are added to the mobile phone. The image English text recognition module is responsible for extracting and recognizing the English text of the image captured by the mobile phone photo or camera, and then the translation module translates the text information into the familiar language of human beings. Then through the retrieval module to these text information in the form of keywords, on the
The algorithm is not effective when the specific letter, nor does it depend on the direction of the is obtained [19], [20]. This method does not depend on a central boundary. Through the above method, a text string isfy certain geometric and spatial characteristics, such as the number of small-scale boundaries. The skeleton must sat- and the outline of a character is represented by a large They believe that the text string is represented by a bound- ing image text using several scale boundaries [4], [17], [18].

They also proposed a new method of find- intervention, so it is only suitable for manual digital camera, background scene with lighting influence, it needs external method can solve the problem of text extraction in complex heuris- knowledge to verify the text region. Although this region search to find the adjacent region, and finally using seed and other pixels in the image is calculated. Using adap- of the text region is obtained, and the distance between the seed and other pixels in the image is calculated. Using adapt- binarization method, connected region analysis, iterative region search to find the adjacent region, and finally using heuris- knowledge to verify the text region. Although this method can solve the problem of text extraction in complex background scene with lighting influence, it needs external intervention, so it is only suitable for manual digital camera, and can not automatically complete the extraction of image text region. Researchers have proposed a new method of find- image text using several scale boundaries [4], [17], [18]. They believe that the text string is represented by a bound- ary model, its centerline is represented by a rough scale, and the outline of a character is represented by a large number of small-scale boundaries. The skeleton must sat- isfy certain geometric and spatial characteristics, such as the restriction that it is perpendicular to or not parallel to the central boundary. Through the above method, a text string is obtained [19], [20]. This method does not depend on a specific letter, nor does it depend on the direction of the text string [21]. The algorithm is not effective when the background is complex. Relevant scholars convert the color image into grayscale image, then through gray stretching to extract edges and binarization, remove noise to get some text areas [22]–[24]. After segmentation by using region splitting and merging algorithm, heuristic knowledge and morphological operation are used, and then the average connected regions are ana- lyzed, and the horizontal and vertical mean values between the marked regions are compared with the first candidate text region. The text region is obtained [25], [26]. Merge the results of the two parts to get the final text area. Although scholars at home and abroad have done a lot of research on text extraction and recognition, and achieved phased results [27]–[29]. However, researchers often select independ- ent databases for different research purposes, so it is difficult to evaluate which method is better and is the best in extracting scene images. Moreover, most of the studies are only aimed at a relatively simple or single background, and the extraction of text with complex image background is still in the initial stage [30], [31].

On the basis of consulting a large number of relevant litera- tures at home and abroad, this paper makes an in-depth study on the characteristics of radial wavelet function, the prop- erties of radial wavelet transform and the characteristics of related dynamic measurement signals, and puts forward some effective solving measures. The correctness of the design idea is proved by numerical calculation, and satisfactory results have been obtained in the application of practical engineer- ing. Specifically, the technical contributions of this paper can be summarized as follows:

First, various signal extraction techniques based on radial wavelet transform modulus Maxima are studied and analyzed in detail, and it is found that these techniques have poor ability to extract weak state signals. And the higher requirement of directivity leads to the pseudo-boundary phenomenon of two-dimensional image extraction results. Through the analy- sis of the correlation denoising method of radial wavelet coefficients, the radial wavelet entropy is introduced into the technical field of signal extraction. The complex Morlet radial wavelet is selected as the wavelet basis function, and a complex Morlet radial wavelet entropy extraction algorithm is proposed, which is suitable for extracting the English text characteristics of weak signal color images.

Secondly, a scene text description method based on Dense-SIFT and spatio-temporal histogram is proposed. Based on the classical visual word bag model, this method uses spatio-temporal histogram to describe the structure information of the text. At the same time, because the spatio-temporal histogram has flexible constraints on the structural characteristics, it can adapt to the changeable shape of the scene text and describe the scene text effectively.

Third, experiments are carried out on a representative benchmark set. The experimental results show that because the spatial histogram can flexibly describe the structural information of the text, the text description method proposed in this paper has strong description ability, can effectively
reflect the characteristics of the text, and has a strong ability to adapt to the distortion in the scene image, and its performance is better than the text description method based on visual word bag model and Ho G. The model compression method based on boundary samples and local extractors proposed in this paper can obtain equivalent compression extractors with fewer pseudo samples, which can effectively improve the efficiency of model compression.

The rest of this article is organized as follows. Section 2 discusses the related work. In section 3, an algorithm of extracting English text from color image by radial wavelet entropy is constructed. In section 4, simulation experiments are carried out, and the experimental results are fully analyzed. Section 5 summarizes the full text.

II. RELATED WORK

A. THE CHARACTERISTICS OF THE TEXT IN THE IMAGE

The text of an image contains rich features. Which of these features are the most useful and how to make use of these features is a key issue to be studied in text location and extraction. The features of the text are complex and meticulous. If you want to locate and extract the text effectively, you need to use a variety of pattern synthesis methods based on features.

The characteristics of the text area (mainly manual text) are mainly considered from the following aspects:

1. Rich color image English text, corners, high intensity and high frequency of periodicity.
2. Most of the text is monochromatic and has obvious contrast with the background, and has a specific color in special cases.
3. The size of the text has a certain range, the distance between the text will not be too large, and a paragraph of text is generally on the same horizontal or vertical line.
4. Image color image English text (subtitle).
5. There is little occlusion in front, most of the text is positive, and the same paragraph will appear in multiple consecutive frames.

These features of the above text can be used to extract the artificial text region, and the scene text usually does not have the above characteristics, because it can appear in different sizes, shapes and colors. Therefore, the research on text region extraction based on scene text has just started, and there are few reports in the literature.

On the feature image obtained by color image English text space mapping and consistent H space mapping, a sliding window with the size of m × n is used to obtain the local data of the image, and the data in the window is extracted from the features. Send it to the extractor to determine whether the corresponding area of the window is a text region. Good features should achieve the goal of strong expression, strong discrimination, stability and easy extraction, so extracting appropriate features will determine the limit performance of recognition, which is the key step of the whole system research.

The easiest way to describe image features is to directly use the pixel gray value of the image. The advantage of this method is that it does not need a specific feature calculation and conversion process, and the feature extraction is fast. However, because the image is generally composed of a large number of pixels, if the gray value of the pixel is directly used as the feature, then the dimension of this feature is very high. For example, if we use a window of 16 × 16, then if we directly use the gray value of the pixel as the feature, the dimension is 256 dimensions. For such a high-dimensional feature, in order to ensure that the recognition learning machine is not too complex, the number of samples must be reduced. This will lead to the decline of learning promotion ability and the ability to extract unknown data. The English text feature discovery and tracking diagram in the color image is shown in Figure 1.

B. MULTIRESOLUTION ANALYSIS

Multi-resolution algorithm comes from the simulation of human eye perception process in computer vision. Generally speaking, multi-resolution structure is a data structure that provides successive compression representation for input image information. The compressed information can be a simple image grayscale, so the successive levels of the multi-resolution structure represent the input image whose resolution decreases step by step. Of course, this structure can also be used to describe some feature information in the image, and these features are represented more and more coarsely at all levels in succession. The formation of the multi-resolution structure is based on the bottom-up calculation of the image, and each layer of the image is formed by some kind of template filtering.

Multi-resolution structure can effectively perform many basic image operations, and it can be used to generate a group of low-pass or band-pass images. Through the interconnection between levels, multi-resolution structure provides the relationship between pixel-level processing and global target-level processing. Two common Multiresolution decomposition structures are pyramid decomposition structure and radial wavelet decomposition structure.

1) BILINEAR INTERPOLATION

The method of bilinear interpolation can be used to decompose an image. Before using bilinear interpolation to reduce the size of an image, a low-pass filter is called to filter the image. This process will reduce the ripple effect caused by the resampling process.

The bilinear interpolation method is to use the linear interpolation method to calculate the value of f (u0 ~ V0) by interpolation according to the gray values of the four adjacent points. Because the bilinear grayscale interpolation method has taken into account the influence of the direct adjacent points of the (u0) point on it, it can generally get a satisfactory interpolation effect.
2) PYRAMID DECOMPOSITION OF IMAGE

The size of the text on the natural scene image varies widely, such as the height of the text on the image provided by ICDAR ranges from 10 to 1200 pixels, some single text accounts for more than 50% of the whole image area, while others are less than 0.1%. At present, almost all text localization algorithms are very sensitive to text size. In order to find text regions of different sizes, this paper uses the method of pyramid decomposition: the image is decomposed into three sub-images of original resolution $1_{\text{max}}$, $1_{\text{max}}^{2}$ and $1_{\text{max}}^{4}$, and the same text region location algorithm is used for each sub-image. Then the location results of each sub-image are fused to generate candidate text regions. Figure 2 shows the effect of pyramid decomposition, in which, a) is the original image, b) is the three sub-images after pyramid decomposition, c) is the extractor corresponding to three sub-images, and the extraction result, d) is the candidate text region generated by image fusion.

3) CONSISTENT HOMOGENEITY

Homogeneity has been successfully applied to image segmentation, because the text region in the image can also be regarded as an independent region that is basically similar within the region, so this paper applies Homogeneity to text region extraction.

Homogeneity is related to the local information extracted from the image, and its value reflects the consistency of a region. We define Homogeneity as a combination of two parts: standard variance and discontinuity of intensity.

In the calculation, the size of the window $m$, $n$ affects the Homogeneity value of the pixel of the point. First of all, the window should be large enough to reflect the local information of the region, but because increasing the size of the window will increase the complexity of the operation, so synthesizing the above two points, the window used in calculating the standard variance is 5, that is, $n = 5$, and the window used in calculating the discontinuity is 3, that is, $m = 3$.

The main characteristics of the text in the image can be summarized as discontinuity and high frequency. Consistent Homogeneity can well reflect the nature of the text in the image, so the method of text extraction and location can be carried out in Homogeneity space.

C. R-CNN SERIES NETWORK

R-CNN series of algorithms, including three network frameworks: Region-based Convolutional Neural Networks (R-CNN), Fast Region-based Convolutional Neural Networks (Fast R-CNN) and Faster Region-based Convolutional Neural Networks (Faster R-CNN). These three models perform target detection tasks on 20 categories of PASCAL VOC data sets, and have good performance. Figure 3 shows the specific structure of these three models.
The original R-CNN model uses the traditional Selective Search (SS) method to process the image, and normalizes the size of the detected 1000 ~ 2000 candidate target regions. Then use the pre-trained convolution neural network to extract the features of the candidate region, and on this basis, through the classification results of support vector machine, predict the category of each candidate region. Finally, the regression device is used to accurately modify the coordinates of the bounding box of each target. The model needs to extract depth features from all candidate regions in turn, so it consumes a lot of time. Moreover, the features used to train the classifier also cause great pressure on the storage space.

In order to solve the above two problems, on the basis of generating 1000 ~ 2000 candidate regions, the convolution neural network is used to extract features from the whole image, and all the candidate region bounding frame coordinates are mapped proportionally to the feature space. Finally, the feature region corresponding to the bounding box is used for category determination and position regression. Compared with R-CNN, this model simplifies the feature extraction process for all candidate regions. At the same time, the ingenious idea of combining category determination and position regression into a multi-task model perfectly solves the problems of training time and space consumption.

Finally, the proposed Faster R-CNN model, on the basis of Fast R-CNN, integrates the Region Proposal Network (RPN) structure for candidate target region extraction, and uses the multi-task model of category determination and location adjustment to solve the problem that traditional detection methods are too time-consuming.

Compared with the R-CNN series, the SSD network chooses to think about the target detection problem from another perspective, that is, to directly extract the original image features, and perform target bounding box coordinate regression and category determination on multiple default positions on the feature map. Subsequent connection of specific convolutional layers makes the feature map size gradually shrink. Finally, the features extracted from different convolutional layers are combined to predict the target object category and the zoom scale of the target bounding box coordinates.

### D. EDGE EXTRACTION ALGORITHM

As we all know, edge is the most basic feature of an image, and the so-called edge refers to the collection of pixels with contrast changes in the surrounding gray intensity, which is not only an important basis for image segmentation, but also an important basis for texture analysis and image recognition. The first step of image analysis and resolution is often edge detection. The edge mainly exists between the target and the target, the target and the background, the target and the region. The ideal edge detection should correctly solve the existence, authenticity and orientation of the edge. For a long time, people have been concerned about the research of this problem. In addition to the commonly used local operators and various improved methods developed on this basis, many new techniques have been proposed, among which LOG, uses Facet model to detect edges, the best edge detector of Canny, statistical filter detection and three-dimensional edge detection with the rise of tomography technology. Although the implementation methods of these operators are different, their purpose is to extract the edge of the image. This paper only discusses the typical Roberts operator, Sobel operator, Canny operator and LOG operator. Finally, a color Robert operator is used to compare the difference of edge extraction from grayscale image and color image. The principles of the first few typical edge extraction operators are not introduced in detail in this paper, only the last color Robert operator is introduced.

The Color Roberts operator firstly extends the Robert operator of grayscale image edge detection to the color image.
edge detection. Instead of simply applying the Robert operator to each component of the pixel color value of the color image, the Euclidean distance is used to comprehensively consider each component of the pixel color value. By using a kind of best edge detection operator LOG operator in grayscale image, the gray edge of the extracted color image is extracted twice, and then the binary edge image is guided to complete the edge extraction of color image.

Figure 4 shows the results of various edge operators. Where, a) is the original graph, b) is Sobel operator, c) is Canny operator, d) is LOG operator, e) is Robert operator and, f) is the result of color Robert operator.

III. AN ALGORITHM FOR EXTRACTING ENGLISH TEXT OF COLOR IMAGE BASED ON RADIAL WAVELET ENTROPY

A. SIGNAL EXTRACTION BASED ON MODULUS MAXIMA OF RADIAL WAVELET TRANSFORM

There may be many points with relatively large gradient amplitudes in the signal, and these points are not all objects that need to be detected in a specific application field, so a screening plan must be determined. The simplest detection criterion is to set the gradient amplitude threshold. Once the signal feature information is extracted, its position can be estimated through multi-scale analysis.

One-dimensional signal can be represented by its amplitude and color image English text, while two-dimensional image signal can be represented by its color image English text and texture characteristics. Gaussian function is a very important smoothing function in color image English text extraction. In general, if $\theta (x)$ is a smoothing function and radial wavelet $\psi (x)$ is the first derivative of $\theta (x)$, then the color image English text of a function $f (x)$ at scale $s$ is defined as the local mutation point of $f (x)$ smoothed by $\theta s (x)$. The modulus Maxima of radial wavelet transform can be used to extract mutation points in multi-scale signals, as shown in Figure 5.

From the definition of transient signal, it is very similar to the English text signal of color image, which has the characteristics of short time, non-stationary, energy concentration and so on. However, due to the high directivity requirements of two-dimensional signals, the extraction methods are relatively complex. Therefore, the current research focus of signal extraction technology is image color image English text extraction. This section discusses the development of signal extraction technology through a detailed analysis of image color image English text extraction technology.

B. ENGLISH TEXT EXTRACTION FROM COLOR IMAGES OF ENGLISH TEXT

Whether a pixel is a modular maximum point can be determined from the structure of the digital image. There are
only eight adjacent points around each pixel, which divide a plane into eight sectors, as shown in Figure 7. The average direction of each sector is represented by a black arrow in the Figure, indicating the discrete gradient direction. Therefore, only 8 directions can be used as gradient directions at each point. Considering the symmetry of gradient directions, only the gradient directions in 0, 1, 2 and 3 sectors need to be considered.

FIGURE 7. The discrete gradient direction of a pixel in an image.

The boundary points in the image generally form a curve, and the curve is usually the boundary of some important structures. The maximum points of each radial wavelet transform modulus are connected to form a maximum curve along the boundary. In the discrete case, the maximum curve is formed by connecting two adjacent boundary points in the image discrete sampling points.

The radial wavelet transform modulus Maxima method can effectively extract the English text of the target color image, but sometimes only want to extract the main outline of the target, and suppress other boundary structures, such as text text. This kind of target is mainly composed of the outline of stepped boundary points, Dirac impact boundary and noise. If the radial wavelet transform modulus Maxima is used to extract the boundary directly, because it can not distinguish the structure of the boundary, all the boundaries will be extracted. In view of this situation, a ladder color image English text extraction method is proposed. Similar to the one-dimensional radial wavelet transform, the attenuation of the two-dimensional radial wavelet transform depends on the regularity of the function f (u,journal v). Here only the case where the Lipschitz index is 0 ≤ α ≤ 1 is considered. If there is a constant K > 0, so that for all (u, v) ∈ Ω ⊆ R²:

\[ |f(u, v) - f(u_1, v_1)| < K(|v - v_1|^2 + |u - u_1|^2) ^{\alpha} \]  (1)

Then the function f is said to be Lipschitz α at the point (U1Magne v1) and is a Lipschitz Ω α above, which is similar to the one-dimensional case. It can also be deduced that if f is uniformly Lipschitz α in a bounded region of R2, if and only if there is a constant A > 0, such that for everything in this region and any scale s > 0, there are:

\[ MWf(s, u, v) < As^{\alpha+1} \]  (2)

Lipschitz assumes that there is an isolated color image English text curve in the image, and the Lipschitz index of the color image English text curve f is α. In some two neighborhood of the English text curve of the color image, the value of, mwf (sdiary u,j.v) can be controlled by the modulus of the radial wavelet transform along the English text curve of the color image.

In the English text extraction of color image, if the same threshold is taken for the whole transformed image, then the local Maxima formed by the English text of the weak fine color image will be filtered out along with the modulus Maxima caused by uneven grayscale, noise and so on. Therefore, the adaptive block method can be used to determine the threshold:

Firstly, the image is divided into many small blocks, and the average value of the modulus Maxima of the radial wavelet transform is calculated in these small blocks. If the average value is less than a certain lower limit, it is considered that there is no English text point of the color image in this region. The points whose modulus is greater than or equal to the average value are output as the English text points of the color image, while the points lower than the average value are filtered out. The design steps of adaptive threshold color image English text extraction method are as follows:

(1) the original image is transformed by radial wavelet transform to generate the modulus family and phase angle family.

(2) the local modulus Maxima points along the phase angle direction are found in the module family, and these local modulus Maxima points are preserved. At the same time, the pixels of other non-local modulus Maxima points are marked as zero, and the English text image of approximate color image is obtained.

(3) the English text image of the approximate color image is scanned by the window, and the threshold T is obtained from the radial wavelet transform coefficient in the window according to the adaptive threshold calculation formula. The point in the image under the window whose modulus Maxima of the radial wavelet transform is greater than T is output as the English text point of the color image; otherwise, the English text point of the non-color image is filtered out.

(4) move the scanning window, scan the English text image of the approximate color image sequentially, and find the English text points of the color image according to step 3 until all the English text points of the color image are obtained. Until a new English text image of a color image is obtained;

(5) repeat steps 1 to 4 until a new English text image of a color image at all scales is obtained.

(6) output color images and English text images at all scales.

Through the comparison of the classical edge detection method and the edge detection method based on wavelet transform modulus maximum, it can be seen that the edge detection algorithm based on wavelet analysis is actually a segmentation technology based on edge detection and edge connection. Different from traditional edge
TABLE 1. Performance comparison of several English text extraction methods for color images.

| Method                                           | Noise suppression ability                                                                 | Ability to extract English text from color images                                      | Amount of calculation | Applicable scope                      |
|--------------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-----------------------|---------------------------------------|
| Roberts                                          | The performance of noise suppression is the worst, and it is impossible to separate English text from noise in color images. | The effect of English text extraction in color image is the worst, and the location of English text in color image is not correct. | Minimum               | No noise pollution, high image definition |
| Sobel                                            | The performance of noise suppression is slightly better.                                  | Color image English                                                                     | Smaller               | No noise pollution                    |
| Laplace-Gaussian                                  | The performance of noise suppression is general, and the English text of color image can be partially extracted. | The English text of the two-pixel wide color image is generated, and the English text direction of the color image cannot be provided. | General               | The contrast is strong and there is a clear outline. |
| A method for extracting Modulus Maxima of Radial Wavelet transform | The anti-noise ability is strong, and the English text of the main color image can be extracted. | It can accurately extract the English text of the color image, and partly provide the direction of the English text of the color image. | A little bigger       | The background is not complex and the noise pollution is general. |
| English text extraction method for stepped Color Image | Strong anti-noise ability, unable to extract detail color image English text               | Only the outline of the target can be extracted, but the details cannot be extracted.   | General               | The contrast is strong and the outline is obvious. |
| Adaptive threshold extraction method              | The effect of noise suppression is the best, and the English text of detail color image can be extracted. | The English text extraction effect of color image is the best, and the English text of weak color image can be extracted. | Maximum              | High extraction accuracy is required, regardless of the amount of calculation. |

detection operators, wavelet operators with multi-resolution analysis capabilities and local mutation detection capabilities are used. Therefore, the modulus local maximum method can be used to detect the fine edge of the signal edge.

However, when the wavelet transform modulus maximum is used for denoising, the signal is greatly affected by the noise at a small scale; at a large scale, the signal will lose some important local singularities. The same problem exists when it is used for signal detection, which leads to the limitation of its detection ability. It is insensitive to weak signals or small amplitude signals, especially when used for signal classification and detection.

C. WEAK SIGNAL EXTRACTION BASED ON RADIAL WAVELET ENTROPY

Generally speaking, the English text of the signal color image has a relatively large amplitude on each scale of the radial wavelet decomposition, and the amplitude of the noise decreases relatively with the increase of the scale. Therefore, by selecting the appropriate radial wavelet basis, the image is processed by radial wavelet analysis to eliminate part of the noise. On the other hand, the radial wavelet decomposition distributes the target signal and noise to each frequency band, because the noise is uncorrelated between each frequency band, and the signal has more energy in each frequency band relative to the noise. Therefore, the correlation denoising method of radial wavelet coefficients can relatively improve the energy of English text of color image and reduce the energy of noise. The combination of the two can achieve the purpose of suppressing noise and highlighting the English text of the color image of the signal at the same time.

In the radial wavelet transform domain of the signal, the amplitude of the radial wavelet transform coefficients represents the intensity of the gray change of the original signal at this resolution, and the points with larger local energy values represent the obvious characteristics of the original signals. Therefore, the energy value of each point can be calculated by the value of radial wavelet transform coefficients. The performance comparison of several English text extraction methods for color images is shown in Table 1.

The difference of time-frequency distribution of different signals is shown by the difference of time-frequency interval energy distribution of different sub-blocks. The theory of radial wavelet entropy is a theory similar to information entropy based on radial wavelet analysis, which can quantitatively describe the characteristics of energy distribution in time-frequency domain. The coefficient matrix of radial wavelet analysis is treated as a probability distribution.
sequence, and the entropy calculated by it reflects the sparse degree of the coefficient matrix.

The concept of radial wavelet entropy is introduced, and the radial wavelet coefficient entropy and radial wavelet correlation entropy with signal length $N$ under scale $j$ are defined as:

$$H_j(W) = -\sum_{n\in N} p_W(j,n) \times \log_2 p_W(j,n)$$ (3)

$$H_j(Cor) = -\sum_{n\in N} p_{Cor}(j,n) \times \log_2 p_{Cor}(j,n)$$ (4)

The radial wavelet transform of the signal has a strong correlation between the radial wavelet coefficients at the corresponding positions of each scale, especially the position of the singularity is neat and has a strong correlation; the radial wavelet coefficients of noise are uniformly distributed and weakly correlated or uncorrelated, especially the radial wavelet transform of noise is mainly concentrated in all levels of small scale. The statistical characteristics of all kinds of noise in the radial wavelet transform domain are similar, and the amplitude of the signal will be greater than that of the noise on a large scale. Therefore, it can be considered that the smaller the entropy of the radial wavelet correlation coefficient is, the more obvious the characteristic of the signal is, and the larger the entropy of the radial wavelet coefficient is, the more obvious the characteristic of noise is. Therefore, the radial wavelet coefficient entropy can be used to determine the noise threshold, and the radial wavelet correlation coefficient entropy can be used to determine the boundary profile.

The theory of radial wavelet entropy is a method to suppress the irrelevant components and realize the accurate location of the signal by using the sparsity of the radial wavelet analysis matrix. Here, the high frequency signal component of each decomposition scale may be regarded as a separate signal source, and the radial wavelet coefficients contained in it are divided into equal intervals, and the radial wavelet entropy of each interval is calculated. The interval with the maximum entropy of radial wavelet coefficients is selected as the detail signal set to estimate the variance of noise.

Assuming that the high-frequency radial wavelet coefficients of layer $j$ are $W_{jmagn}$, the correlation coefficients of radial wavelets are $Cor_{jmagn}$, and the number of sampling points is $N$, the radial wavelet coefficients of these sampling points are divided into $m$ equal parts, then the radial wavelet coefficients of the $k$-th sub-interval are $W_{jmagnet k}$ and the radial wavelet correlation coefficients are $Cor_{jscene k}$. From the characteristics of orthogonal radial wavelet transform, it can be known that in a certain time window, the total power of the signal is equal to the sum of the power of each component.

$$E = \sum_{j,k} |W(j,k)|^2 = \sum_j E_j$$ (5)

The extraction of the English text points of the secondary important color image of the signal can be calculated and determined in turn in the remaining radial wavelet coefficients. The principle block diagram of weak signal extraction based on radial wavelet entropy is shown in Figure 8 shows the extraction process of one-dimensional signal and two-dimensional image, respectively.

Complex Morlet radial wavelet is a kind of single-frequency complex sinusoidal modulated Gaussian wave, and it is also the most commonly used complex-valued radial wavelet. It has good locality in both time domain and frequency domain, and the spatial corresponding characteristic is even better than that of Mexican hat radial wavelet. The reason why this paper chooses the non-orthogonal complex Morlet radial wavelet as the radial wavelet basis function to extract the English text of transient signal or image color image is mainly for the following reasons:

1. The amplitudes of the real and imaginary parts of the complex Morlet radial wavelet are both exponentially attenuated harmonic vibration signals, which are consistent with the free response signals of the dynamic system.

2. The complex Morlet radial wavelet has a single frequency, and if the analyzed signal is highly correlated with the radial wavelet at a certain scale, the corresponding radial wavelet frequency represents the natural frequency of the dynamic system in which the response signal is generated.

3. Dyadic orthogonal radial wavelet has good signal reconstruction ability, but its scale is binary discrete, so the frequency resolution of this kind of radial wavelet transform will be limited, while using non-orthogonal Morlet radial wavelet with no scale function to carry out signal continuous radial wavelet transform can achieve arbitrarily high resolution in time domain or frequency domain. This advantage has a good effect on the English text recognition of transient signals or image color images.

4. The complex radial wavelet has good directivity and shift invariance and provides phase information, and the redundant information in the radial wavelet coefficients can also be controlled by its construction method, which is very suitable for image decomposition.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL DATA

In the experiments in this section, three popular benchmark data sets for scene text recognition are used, including c74kmag S VTC-Char and ICDAR2003.

ICDAR2003 dataset is the most popular benchmark dataset for text recognition algorithms in test scenarios. This data set was originally established for the text recognition competition part of the document analysis and processing conference. The dataset contains 62 types of image samples with a total of 11615 English texts (A, Z, a, z) and Arabic numerals (0; 9). The disadvantage of this data set is that the number of other samples extracted from a small number of data sets is very rare.

SVT data set is a scene text image data set established for the word calibration task in the scene image, which contains
Y. Wang: Extraction Algorithm of English Text Information From Color Images Based on Radial Wavelet Transform

FIGURE 8. Weak signal extraction based on Radial Wavelet Entropy and signal extraction process of two-stage Radial Wavelet transform.

Based on the SVT dataset, the location information of the text is added and the new dataset is named SVT-char, which contains 52 categories (A VT a char,) with a total of 4000 text image samples.

C74k data set is also a benchmark data set often used in scene text recognition algorithm testing, which contains English text, Arabic numerals and Kannada text image samples. In addition to scene text samples, a set of artificial text samples can be generated using the built-in fonts in the Windows system, with a total of 1016 image samples in 4 different font styles and 254 fonts for each type of text. It is worth noting that in addition to the more commonly used sample data, 5000 text samples with serious occlusion and blur are provided in c74k. For ease of expression, in the following experiments, the better quality scene text samples in C74k are recorded as c74k-Good, and the worse quality samples are recorded as c74k-Bad. According to the existing literature, due to the poor quality of the above samples, there is no method to achieve better recognition results on c74k-Bad data. This section compares the proposed method with the mainstream method on c74k-Bad data, and the results are encouraging.

B. EXPERIMENTAL PARAMETER SETTING

1) PRETREATMENT

The experiments were conducted separately on each benchmark set. In the course of the experiment, the data set is randomly divided into training set, verification set and test set. The number of samples extracted from SVT-char and ICDAR 2003 datasets is relatively small, which leads to the imbalance of samples during extractor training. For this reason, in the course of the experiment, the sampling proportion of this category is increased accordingly, and salt and pepper noise, Poisson noise, Gaussian noise and speckle noise are randomly added to the repeated samples to enrich the content of the training samples. As a small number of other samples are extracted, the above methods can only weaken the problem of sample imbalance to a certain extent.

There are differences between image samples in the benchmark data set. For example, the samples in the data set include both grayscale images and color images, and the sample sizes are not consistent. Therefore, it is necessary to preprocess the samples. Specifically, the color image samples are converted into grayscale images, and all samples are stretched or downsampled at the same time, so that the resolution of all images is adjusted to 128 × 128. In addition, in view of the low contrast of some samples, the histogram equalization is carried out to improve the contrast of the above samples.

In the course of the experiment, Dense-SIFT is used to extract the local descriptors in the sample. The specific parameters of the extraction process of Dense-SIFT descriptors are set as follows: the sampling step is set to 3, and the pre-smoothing scale parameters are 4, 6, 7, 10. In the learning process of the visual dictionary, the K-means algorithm is used to obtain visual words. In the clustering process, the number of category centers is set to 60, even if the 60 × k dimensional feature vector is used to describe the sample, where k is the number of sub-regions.

2) BASE EXTRACTOR TRAINING

In the extractor design training stage, the “one-to-one” strategy is used to decompose multi-class text recognition problems into two kinds of problems. A large number of heterogeneous base extractors are trained for each extraction sub-problem, including KNN, ANN, TREE and radial wavelet entropy extractors. For each class of extractors, the diversity of extractors is ensured by using different training parameters (ANN and radial wavelet entropy) and adjusting training samples (TREE and KNN). Specifically, for the
radial wavelet entropy extractor, the value of parameter $C$ is adjusted so that it varies from 0.01 to 1000, and 2000 extractors are obtained. Similarly, for ANN extractors, 2000 ANN extractors are obtained through different hidden nodes, neural network layers and iterations. For KNN extractor and TREE extractor, using the method of Bagging, 2000 extractors are also constructed for each type of extractor. Because the performance of the extractor is not tested in the above process, the effect of the extractor is good and bad, but because of the large distribution range of training parameters, it can cover the extractor with the best performance.

3) EXTRACTOR INTEGRATION
In the extractor integration phase, the output of each base extractor is calculated on the verification set and the performance of the base extractor is calculated. In the process of the performance of the statistical base extractor, the accuracy of all samples is not counted, but the accuracy of each category is calculated and averaged separately.

In the next stage, the base extractors are sorted according to their performance, and the first 30 base extractors are selected as initialized integrated extractors. The recognition of the integrated extractor adopts an unweighted voting method, which is mainly due to the fact that the algorithm is put back in the selection process, and the strategy essentially implies the weighted operation of the extractor. Then the base extractor is selected from the base extractor collection and added to the integration, and the upper limit of the number of extractors is set to 100.

4) MODEL COMPRESSION
The equivalent extractor is obtained by compressing each of the two types of extractors. Specifically, boundary samples and MUNGE pseudo-samples are generated for each of the two types of extractors at a ratio of 0.3 and 0.7 respectively. Then use the MUNGE pseudo-samples to train an ANN extractor, the extractor includes two hidden layers, for each two kinds of problem training, in addition to the input layer and output layer nodes are fixed, adjust the number of hidden layer nodes to get the optimal results. Then the self-organizing feature map network is used to cluster the BDS samples, and k centers are obtained. Because of the serious imbalance of samples in each clustering center, samples are randomly selected from the categories with a large number of samples to find the nearest neighbor of the sample in the BDS sample, add it to the category center, and repeat the above process until the balance is reached. In order to ensure the efficiency of the local extractor, the algorithm trains a linear radial wavelet entropy extractor for each clustering center.

C. DESCRIPTION EXPERIMENT BASED ON SPATIO-TEMPORAL HISTOGRAM
Firstly, the spatio-temporal histogram method is compared with a group of typical methods based on local description, including the method based on visual word bag model, the method based on nearest neighbor and the method based on Constellation model. It should be noted that at this stage, we do not use the integrated extractor, but use the radial wavelet entropy of the Chi2 kernel as the text extractor.

The method based on the nearest neighbor of the template uses the number of local features matched in the two text images as a measure of similarity, and finds the nearest neighbor in the template image to realize the scene text recognition. Specifically, the method generates a set of template images for each type of text, and the category of the sample to be recognized is determined by the template image of its nearest neighbor. This method uses MUNGE method to match local features and actually describes the structure information of the text.

The method based on Constellation model is a partial production model, which was first used for target extraction and object class recognition. We implement this method and use this model to realize the scene text recognition. The model uses a set of local features on the text to describe the text. Among them, the mixed Gaussian model is used to describe the apparent information of local features and the position relationship between features.

Visual word bag model is one of the most popular image description methods in the field of computer vision. The implementation of this method is relatively simple, and the effect is good, so it has been widely used.

In the above methods, the structure of the model is very different. In order to ensure the objectivity of the comparative experiment as much as possible, some parameters are set as follows: (1) based on the nearest neighbor method, each category generates 15 templates; (2) in the methods based on Constellation model, classical Bo Ws and spatial histogram, each class uses 15 training samples to train the extractor.

Then, this method is compared with the scene text recognition method based on global features, including the method based on random syncope and the method based on radial wavelet entropy. The random syncope-based method is a part of the end-to-end scene text recognition system, which uses the Ho G feature to describe the sample, while the random syncope-based extractor is used for recognition. The method based on radial wavelet entropy, as a part of text-based image retrieval system, uses HOG features based on RBF kernel and linear kernel radial wavelet entropy extractor to realize scene text recognition.

C74k data. Through the analysis of the error results, it is found that on average, about 8% of the errors are caused by similar texts such as ‘p’ and ‘P’, ‘z’ and ‘Z’. It is worth noting that the recognition accuracy of RBF and Linear, is based on the optimal selection of training samples and features. In this method, the samples are selected randomly, and the average accuracy is obtained many times, in fact, the influence of sample selection is eliminated. In this case, the method proposed in this paper still has advantages in performance, which fully proves that the visual word bag model based on spatio-temporal histogram can effectively describe the scene text. A fixed number of training samples are used in the above
experiments, and the effect of the number of training samples on the performance of the extractor is investigated below.

In the experiment, the method in this paper is compared with the method in Linear on three different data sets. In the process of comparison, the number of training samples is changed to investigate the impact of the change on the performance of the algorithm. The results are shown in Figure 9.

From the experimental results, it can also be found that, without optimizing the allocation of training samples, the accuracy of the proposed method on all data sets is significantly higher than that of the method based on HOG, especially when the number of samples is very small. The main reasons for this phenomenon are as follows:

First of all, because the local feature description describes the local structure of the image, which is more stable than the whole text image, it is robust to the degradation and distortion in the scene image and helps to improve the accuracy of recognition.

Secondly, the spatial histogram introduces structural information, which is of great significance for text recognition with strong structural information.

Thirdly, the description of spatial histogram is relatively flexible, so it can adapt to the large differences within the text class of the scene.

D. INTEGRATED LEARNING AND MODEL COMPRESSION EXPERIMENT BASED ON BOUNDARY SAMPLES

From the perspective of pattern classification method research, improvements can be made from two aspects of character model and classifier to improve the accuracy of scene character recognition:

(1) The morphology of scene characters is quite variable, and the differences within the category are relatively large. The ideal character model should be relatively simple and invariant to changes in colors, fonts, scales and possible image distortions. At the same time, the descriptive ability of the model should be relatively strong, able to fully reflect the characteristics of the characters.

(2) The number of categories in scene character recognition tasks is usually relatively large, and there are many variants of each type of character. The collected samples cannot cover all character variants. In order to improve the recognition accuracy, it is necessary to improve the generalization ability of the classifier as much as possible and enhance the ability to process unknown samples.

First of all, this section tests the model compression method based on boundary samples. Specifically, six base extraction problems are selected from the English text recognition task as the test object, and the influence of the parameters is analyzed by observing the performance of the algorithm on the above-mentioned base extractor. The integrated learning method is used to train the integrated extractor for each of the above extraction problems. After the completion of the training, the extraction accuracy of the integrated extractor and the optimal base extractor in each extraction problem was counted.

On the basis of the training of the base extractor and the integrated extractor, we first test the effect of model compression using MUNGE-BS and MUNGE methods on the error rate of the extraction problem under the premise of different sample numbers. Specifically, the proportion of boundary samples is 30%, the number of local extractors is k = 10, the total number of samples changes from 10 times the number of training samples to 100 times, and the step size is 10. The experimental results are shown in Figure 10. From the results, it is easy to find that the model compression method based on boundary samples requires less pseudo-samples and higher accuracy under the same conditions. But interestingly, it is possible that the performance of the compression model is slightly better than that of the target extractor, but this will not happen because the extractor obtained by using the model compression method of boundary samples is closer to the target extractor.

Let’s test the effect of the number of local extractors k on the model compression method. The boundary samples are generated for each problem, and the algorithm is used to generate pseudo samples, in which the boundary samples account for 30%, and the total number of samples is 30 times the number of training samples. Using the above data, the compression model is trained and the value of k is changed, and the effect of the change on the error rate is observed. The result is shown in Figure 11.

From the results, it is found that the number of local radial wavelet entropy helps to reduce the error rate, but the more
the number of radial wavelet entropy, the better. In some problems, the error rate does not decrease with the increase of radial wavelet entropy. On the contrary, the error rate increases.

In order to further verify the performance of the algorithm, the above methods are tested on the dataset used in the previous section. In particular, this method and the comparison method are tested on the c74k-bad dataset. Due to the serious ambiguity and degradation of the samples in the dataset, there is no method to achieve ideal results on the dataset. The sample in this dataset is shown in Figure 12. The results of this method are compared with those of other methods. The results of the comparative test are shown in Figure 13.

Through the experimental results of this section, it can be found that the improvement of accuracy brought by this method on this data set is greater than that on other data sets. This shows that the description method and integrated learning method based on spatio-temporal histogram and visual word bag model can solve the problems of blurred scene text and image degradation to a great extent.

The experiment in this section proves that the ensemble learning method can further improve the extraction accuracy on the basis of the base extractor, mainly because the
ensemble learning method improves the generalization ability of the extractor. This is of great significance for scene texts with large intra-class differences. The essence of the model compression process is to detect the extraction interface of the known extractor by generating pseudo samples and use a new extractor to describe the detected extraction interface. The experimental results of this section show that the boundary samples can promote the above process and can achieve a more accurate description on the premise of using fewer pseudo samples. This is mainly due to the fact that the local extractor based on boundary sample training can fit the extraction interface more accurately in the local space. Therefore, the above experiments further prove that the model compression method based on boundary samples can effectively construct an extractor equivalent to the target extractor, with only a small performance loss while improving the computational efficiency.

V. CONCLUSION

In this paper, through the detailed study and analysis of various signal extraction techniques based on the modulus Maxima of radial wavelet transform, it is found that these techniques have poor ability to extract weak-state signals, and the higher requirements for directionality often lead to pseudo-boundary phenomena in two-dimensional image extraction results. On this basis, the radial wavelet entropy is introduced into the signal extraction field, and the complex Morlet radial wavelet is selected as the radial wavelet basis function. A complex Morlet radial wavelet entropy extraction algorithm is proposed, which is suitable for extracting the English text characteristics of weak signal color images. Using the multi-scale and multi-resolution property of radial wavelet analysis, that is, it has good locality in time-frequency domain at the same time, we have found a more effective research method of signal extraction technology. Based on the coefficient characteristics of different modulus Maxima of signal and noise in radial wavelet transform domain, not only the features of signal and noise in multi-scale resolution space can be extracted, but also the signal features can be extracted effectively according to the different propagation characteristics of radial wavelet transform modulus Maxima. This not only avoids matrix operation and reduces the amount of computation, but also improves the signal-to-noise ratio gain, maintains a good resolution to the details of the signal, and is insensitive to the form of the signal. The scene text has the characteristics of many categories and great differences within the class. The traditional OCR method is difficult to achieve ideal results in the scene text recognition task. In this paper, the problem of scene text recognition is studied, and a scene text extraction method based on integrated learning and model compression is proposed. The ensemble learning method can significantly improve the generalization ability of the extractor, but the speed of the integrated extractor is often slow and takes up more space. In order to solve this problem, this paper proposes a model compression method based on boundary samples and local extractors, which can significantly reduce the number of pseudo-samples needed and efficiently compress the integrated extractor into a more concise compression extractor. In addition, this paper proposes a text feature based on local features and spatio-temporal histogram. The experimental results show that the combination of the two methods can significantly improve the accuracy of scene text recognition.

REFERENCES

[1] E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, Apr. 2017.
[2] Y. Zhu, C. Yao, and X. Bai, “Scene text detection and recognition: Recent advances and future trends,” Frontiers Comput. Sci., vol. 10, no. 1, pp. 19–36, Feb. 2016.
[3] Y. Zhu, M. Liao, M. Yang, and W. Liu, “Cascaded segmentation-detection networks for text-based traffic sign detection,” IEEE Trans. Intell. Transp. Syst., vol. 19, no. 1, pp. 209–219, Jan. 2018.
[4] T. He, W. Huang, Y. Qiao, and J. Yao, “Text-attentional convolutional neural network for scene text detection,” IEEE Trans. Image Process., vol. 25, no. 6, pp. 2529–2541, Jun. 2016.
[5] F. Okura, K. Vanhoey, A. Bousseau, A. A. Efros, and G. Drettakis, “Unifying color and texture transfer for predictive appearance manipulation,” Comput. Graph. Forum, vol. 34, no. 4, pp. 53–65, Jul. 2015.
[6] C. Lia, J. Cao, W. Ranaweera, I. Alhashim, P. Tan, A. Sheffer, and H. Zhang, “Legible compact calligrams,” ACM Trans. Graph., vol. 35, no. 4, pp. 122–112–12-12, Jul. 2016.
[7] S. Zhang, X. Gao, N. Wang, and J. Li, “Robust face sketch style synthesis,” IEEE Trans. Image Process., vol. 25, no. 1, pp. 220–232, Jan. 2016.
[8] C. Yao, X. Bai, and W. Liu, “A unified framework for multi-oriented text detection and recognition,” IEEE Trans. Image Process., vol. 23, no. 11, pp. 4737–4749, Nov. 2014.
[9] O. Le Meur, M. Elhrouci, and C. Guilleminot, “Hierarchical super-resolution-based inpainting,” IEEE Trans. Image Process., vol. 22, no. 10, pp. 3779–3790, Oct. 2013.
[10] B. Shi, X. Bai, and C. Yao, “An End-to-End trainable neural network for image-based sequence recognition and its application to scene text recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 11, pp. 2298–2304, Nov. 2017.
[11] Y. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “PatchMatch: A randomized correspondence algorithm for structural image editing,” ACM Trans. Graph., vol. 28, no. 3, pp. 341–352, Aug. 2009.
[12] E. Carlinet and T. Géraud, “A comparative review of component tree computation algorithms,” IEEE Trans. Image. Process., vol. 23, no. 9, pp. 3885–3895, Sep. 2014.
[13] P. Shivakumara, T. Quan Phan, and C. Lim Tan, “A Laplacian approach to multi-oriented text detection in video,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 2, pp. 412–419, Feb. 2011.
[14] H. Q. Phan, H. Fu, and A. B. Chan, “FlexyFont: Learning transcribing rules for flexible typeface synthesis,” Comput. Graph. Forum, vol. 34, no. 7, pp. 245–256, Oct. 2015.
[15] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, “Drawing and recognizing Chinese characters with recurrent neural network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 849–862, Apr. 2018.
[16] M. Maja and D. G. Lowe, “Scalable nearest neighbor algorithms for high dimensional data,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 11, pp. 2272–2240, Nov. 2014.
[17] X.-C. Yin, Y. Yin, K. Huang, and H.-W. Hao, “Robust text detection in natural scene images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 5, pp. 970–983, May 2014.
[18] J. Gaty, A. S. Ecker, T. and M. Bethge, “Texture synthesis using convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., vol. 07, no. 1, pp. 262–270.
[19] Y. Chang, S. Saito, and M. Nakajima, “Example-based color transformation of image and video using basic color categories,” IEEE Trans. Image Process., vol. 16, no. 2, pp. 329–336, Feb. 2007.
[20] M. Liao, B. Shi, and X. Bai, “TextBoxes++: A single-shot oriented scene text detector,” IEEE Trans. Image Process., vol. 27, no. 8, pp. 3676–3690, Aug. 2018.
[21] Y. Wexler, E. Shechtman, and M. Irani, “Space-time completion of video,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 3, pp. 463–476, Mar. 2007.

[22] X.-C. Yin, Z.-Y. Zuo, S. Tian, and C.-L. Liu, “Text detection, tracking and recognition in video: A comprehensive survey,” IEEE Trans. Image Process., vol. 25, no. 6, pp. 2752–2773, Jun. 2016.

[23] C. Yi and Y. Tian, “Text string detection from natural scenes by structure-based partition and grouping,” IEEE Trans. Image Process., vol. 20, no. 9, pp. 2594–2605, Sep. 2011.

[24] Y. Shih, S. Paris, F. Durand, and W. T. Freeman, “Data-driven hallucination of different times of day from a single outdoor photo,” ACM Trans. Graph., vol. 32, no. 6, pp. 2504–2507, 2013.

[25] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, “Arbitrary-oriented scene text detection via rotation proposals,” IEEE Trans. Multimedia, vol. 20, no. 11, pp. 3111–3122, Nov. 2018.

[26] C. Yi and Y. Tian, “Scene text recognition in mobile applications by character descriptor and structure configuration,” IEEE Trans. Image Process., vol. 23, no. 7, pp. 2972–2982, Jul. 2014.

[27] Y.-W. Tai, J. Jia, and C.-K. Tang, “Soft color segmentation and its applications,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 9, pp. 1520–1537, Sep. 2007.

[28] Q. Ye and D. Doermann, “Text detection and recognition in imagery: A survey,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 7, pp. 1480–1500, Jul. 2015.

[29] M. Elad and P. Milanfar, “Style transfer via texture synthesis,” IEEE Trans. Image Process., vol. 26, no. 5, pp. 2338–2351, May 2017.

[30] C. Barnes, F.-L. Zhang, L. Lou. X. Wu, and S.-M. Hu, “PatchTable: Efficient patch queries for large datasets and applications,” ACM Trans. Graph., vol. 34, no. 4, p. 97, 2015.

[31] S. Zhang, X. Gao, N. Wang, J. Li, and M. Zhang, “Face sketch synthesis via sparse representation-based greedy search,” IEEE Trans. Image Process., vol. 24, no. 8, pp. 2466–2477, Aug. 2015.

Yaqin Wang was born in Shaanxi, China, in 1973. She received the master’s degree from Inner Mongolia Normal University, in 2009. From 2000 to 2008, she was working with Ordos Educational College, Inner Mongolia. She is currently with Ordos Vocational College. She has published a series of research articles, eight of which have been indexed by CNKI. Her research interest includes second language acquisition.

* * *