A Novel Vision-Based Adaptive Scanning for the Compression of Remote Sensing Images

Cuiping Shi, *Student Member, IEEE*, Junping Zhang, *Member, IEEE*, and Ye Zhang, *Member, IEEE*

***Abstract*—Most of the compression methods for remote sensing images are often designed under the guidance of mean square error. However, for the vision-related applications, high peak-signal-to-noise ratio (PSNR) does not mean good visual quality. On the other hand, existing compression methods that considering the human visual system (HVS) are usually designed for natural images, without taking the unique characteristics of remote sensing images into account. Focusing on this problem, we present a novel HVS-based adaptive scanning (HAS) scheme for the compression of remote sensing images. First, after the wavelet transform, a retina-based visual sensitivity model is established, and then, the visual weighting mask is generated. Second, for the weighted transformed image, an adaptive scanning method is proposed, which provides different scanning orders among subbands and within a subband, respectively. The former focuses on organizing the codestream according to the importance of weighted subbands, and the latter aims at preserving the direction information of an image as much as possible. Finally, the binary tree codec is utilized. Experimental results show that, as compared with other scan-based compression methods, the proposed HAS-based compression method can provide better visual quality, which makes it more desirable in vision-related applications for remote sensing images.**

***Index Terms*—Adaptive scanning, embedded coding, human visual system (HVS), remote sensing image compression.**

I. Introduction

Along with the increasing demand for remote sensing data, the sensor technology has been developed to improve the spatial and spectral resolution, which brings great convenience to the application of remote sensing images. On the other hand, abundant information is at the cost of the huge data. At present, the latest generation spaceborne sensors can continuously produce massive quantities of high-dimensional remote sensing images at a rate of several terabytes per day, which brings a great challenge to data storage and transmission [1], [2]. These contradictions can be alleviated by exploiting some traditional compression technologies, such as EZW [3], SPIHT [4], SPECK [5], JPEG2000 [6], or some improved versions of them [7]–[13]. Usually, these compression methods tend to be measured in the sense of the mean square error (MSE), i.e., at the same conditions, the compression method that can provide smaller MSE is considered to be a better method. However, a reconstructed image with smaller MSE does not mean that it is suitable for all applications. In fact, the evaluation of a compression method mainly depends on the application. With the popularity of remote sensing images, many applications require that the remote sensing image can be browsed online. One of the research hotspots, i.e., Digital Earth, also needs a large number of remote sensing images with good visual quality to support it [14]. In this case, a compression method designed with considering human vision is more favorable. The human visual system (HVS) is a complex system, which has been proven to be inconsistent with MSE [15]. Therefore, it is necessary to study a compression method for remote sensing image from the perspective of the human visual mechanism.

Human perception has been and still is the focus of many image coding studies [15]. There are several ways to incorporate human perception into image coding schemes. One way is discrete cosine transform (DCT)-oriented approaches, such as in [16]–[18], and another is discrete wavelet transform (DWT)-oriented approaches, such as in [19]–[22]. In addition, some coding schemes are designed with considering the just-noticeable difference (JND) model, which exploits the fact that some distortions of an image are not perceptible by a human observer, such as in [23] and [24]. In addition, some vision-related compression methods are designed based on JPEG2000, such as in [25] and [26]. Recently, some HVS-based coding schemes have been designed from the perspective of information theory. For example, Niu *et al.* [27] proposed a perceptual coding strategy, which aims at preserving the scale-invariant second-order statistics of natural images to guarantee the perceptual quality, and the authors in [28] developed some new models, with regard to the HVS, and deduced a new limit that promises theoretically achievable data reduction ratios with no perceptual loss in typical scenarios. However, all these approaches are designed for natural images and do not take the characteristics of remote sensing images into consideration.

For a natural image, after the DWT, a compact representation usually can be obtained, which helps to obtain a good coding performance. However, as compared with a natural image,

Manuscript received December 6, 2014; revised May 29, 2015 and August 4, 2015; accepted September 6, 2015. Date of publication October 6, 2015; date of current version February 24, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61271348 and Grant 41301455, in part by the 863 High Technology Program of China under Grant 2012AA12A405, and in part by the Prophase Cultivation Project of Technology Achievements Industrialization of Universities in Heilongjiang Province of China under Grant 1254CGZH04.

C. Shi is with the Department of Information Engineering, Harbin Institute of Technology, Harbin 150001, China, and also with the Department of Communication Engineering, Qiqihar University, Qiqihar 161000, China (e-mail: scp1980@126.com).

J. Zhang and Y. Zhang are with the Department of Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: zhangjp@hit.edu.cn; zhye@hit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2015.2478145

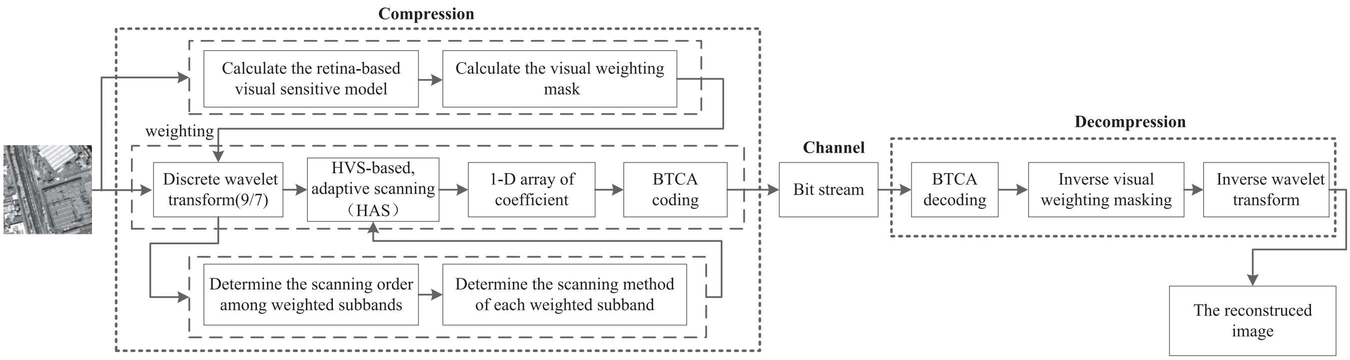


Fig. 1. Overall framework of the proposed HAS-based compression method.

a remote sensing image has its own unique characteristics. It usually contains a great number of ground objects, which makes the information of details abundant, such as geometric information, edge and texture information, and even the outline of some small targets. Therefore, the coefficients of high-frequency subbands are usually still very large, which becomes a heavy burden for compression.

In recent years, some compression schemes that are specifically designed for remote sensing images have been proposed, such as in [29]–[32]. These compression schemes compress remote sensing images from several aspects, such as oriented wavelet transform or sparse representation. Since remote sensing images are often captured by sensors in a push-broom fashion and are quite large, thus, the scan-based compression approach is also very desirable. Kulkarni *et al.* [33] present a scan-based method that can use JPEG2000 with incrementally acquired data. However, the high coding performance of JPEG2000 is at the cost of high complexity. The Consultative Committee for Space Data Systems (CCSDS) published a recommended standard for the onboard image compression. The standard of CCSDS is a scan-based algorithm, but it does not allow the interactive decoding. Furthermore, the level of DWT of it is fixed to 3. In 2009, Vílchez *et al.* [34] presented some prominent extensions to the CCSDS, which allowed any number of wavelet decomposition levels and supported several forms of decoding. However, all these approaches are based on fixed scanning, which does not take the content of an image into consideration.

In 2012, the state-of-the-art compression method based on scanning for remote sensing images, which was named binary tree coding adaptively (BTCA), was proposed [35]. It can improve the coding performance significantly. Although the process of BTCA is somewhat related to the content of the image, it still adopts a fixed scanning way to scan the transformed image before establishing a binary tree. As we know, different images are of different contents. The content of an image can be described from different aspects, such as color, texture, shape, and so on. In this paper, the content of an image is defined from the perspective of energy. The reason is that different images are of different directional information, and the amount of directional information can be reflected from the energy distribution of the transformed image. If the subbands with larger energy can be scanned in priority, at the same bit rate, more directional information can be preserved. Therefore, from the perspective of scanning, an adaptive scanning method is more helpful to improve the coding performance.

The remainder of this paper is organized as follows: In Section II, a retina-based visual sensitivity model is analyzed, and the importance weighting mask is generated. Then, we give a detailed description of the proposed HVS-based adaptive scanning (HAS) scheme for remote sensing images. In Section III, the binary tree codec is introduced. Section IV gives some quality evaluation indexes used in this paper. In Section V, we present some numerical experiments and prove the high validity of the proposed method. Finally, the conclusion and discussion is provided in Section VI.

II. Proposed HAS Method

In this paper, a novel HAS approach for the compression of remote sensing images is proposed. The whole compression process can be viewed as two stages in cascade. The first stage is to generate an importance weighting mask according to the human visual characteristics, which will help to guarantee that the bits with greater contribution to the visual quality of image can be scanned in priority. The second stage focuses on designing different scanning orders among subbands and within a subband, respectively. Finally, the binary tree codec is exploited to encode the 1-D coefficient sequence. The overall framework of the proposed HAS-based compression method is shown in Fig. 1.

A. Retina-Based Visual Sensitivity Model

The human eye uses the retina to collect and process visual information [15]. In the human retina, the spatial distribution of photoreceptors is nonuniform, with the highest density at the fovea. This density rapidly decreases with distance from that area. Hence, the local visual frequency bandwidth also falls away [45]. Usually, the visual frequency is used to describe the contrast sensitivity function (CSF), which is utilized to characterize the varying sensitivity of the visual system to 2-D spatial frequencies [46]. The human eye cannot perceive spatial frequencies beyond a given cutoff frequency, i.e., it is not necessary to preserve the information of very high spatial frequency of an image from the perspective of HVS. Therefore, the characteristics of the retina must be considered, if we want to

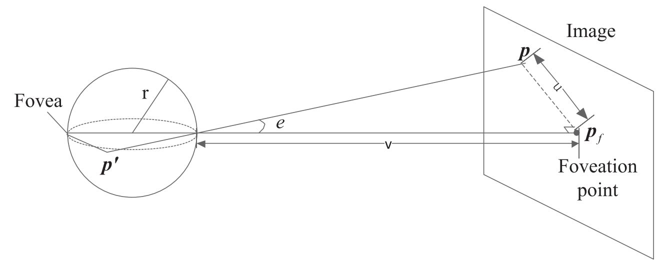


Fig. 2. Mapping model of the fovea and the viewing distance.

improve the visual effect of the reconstructed image at the given bit rate.

For the retina, vision is more sensitive at the fovea than at other portions of the retina. In other words, the spatial resolution of the HVS is the highest at the fovea, and it decreases rapidly with increasing eccentricity. Based on this fact, it is possible to remove considerable visual information redundancy from the peripheral regions and still reconstruct a perceptually good quality image [36]. For a given image, the visual sensitivity model provides the error sensitivity of each position. The less sensitive to the error, the more information can be removed from this position. In [36], Wang and Bovik established the mapping model of the fovea and the viewing distance; it is shown in Fig. 2.

In Fig. 2, $\boldsymbol{p}$ represents any point in the image, and $\boldsymbol{p}'$ is the projection onto the retina of point $\boldsymbol{p}$. $\boldsymbol{p}_f$ represents the point we are staring at, i.e., the foveation point, and the fovea is the projection onto the retina of point $\boldsymbol{p}_f$. Here, we assume that the image plane is perpendicular to the line that passes through the fovea and the foveation point $\boldsymbol{p}_f$. From the mapping model, it can be seen that a circle in the image plane centered at $\boldsymbol{p}_f$ will be projected to a circle in the retina centered at the fovea.

In [37], Geisler and Perry have pointed out that the contrast threshold is a function of retinal eccentricity. For a retinal eccentricity, the contrast threshold function can calculate the maximum frequency that can be perceived by the human eye. The contrast threshold function can be obtained by fitting the experimental data. For an image, the contrast threshold function in the spatial domain can be represented as

$$CT(f,e) = CT_0 \exp\left(\alpha f \frac{e+e_2}{e_2}\right). \tag{1}$$

Here, f represents the spatial frequency (in cycles/degrees); e represents the retinal eccentricity (in degrees). CT_0 represents the minimal contrast threshold. α is the spatial frequency decay constant. e_2 represents the half-resolution eccentricity constant. $CT(f,e)$ represents the visible contrast threshold, which is a function of f and e. For the model (1), the best fitting parameter values given in [37] are $CT_0 = 1/64$, $\alpha = 0.106$, $e_2 = 2.3$.

For a given eccentricity e, (1) can be used to find its critical frequency, namely, cutoff frequency f_c, which means that any higher frequency component beyond it is invisible. f_c can be obtained by setting $CT = 1$, i.e.,

$$f_c = \frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\alpha(e+e_2)} \text{ cycles/degree.} \tag{2}$$

Obviously, the cutting frequency f_c only depends on the eccentricity e. We assume that the width of an observed image is N pixels, and the foveation point $\boldsymbol{p}_f = (p_x^f, p_y^f)^T$ (pixels) and the viewing distance v (measured in image width) from the eye to the image are known. The distance from point $\boldsymbol{p}$ to $\boldsymbol{p}_f$ is then $d(\boldsymbol{p}) = \|\boldsymbol{p} - \boldsymbol{p}_f\|_2$ (measured in pixels). The distance u (measured in image width) from point $\boldsymbol{p}$ to $\boldsymbol{p}_f$ is then $u = d(\boldsymbol{p})/N$. The eccentricity is given by

$$e(v,\boldsymbol{p}) = \tan^{-1}\left(\frac{d(\boldsymbol{p})}{Nv}\right). \tag{3}$$

It can be seen that, from (2) and (3), for the given viewing distance, the cutting frequency is a function of pixel position. On the other hand, the maximum perceived resolution is limited by the display resolution r, i.e.,

$$r = \frac{\pi N v}{180} \text{ pixels/degree.} \tag{4}$$

According to the sampling theorem, the highest frequency can be represented without aliasing by the display, i.e., the display Nyquist frequency is

$$f_d = \frac{r}{2} = \frac{\pi N v}{360} \text{ cycles/degree.} \tag{5}$$

Based on (2) and (5), the final cutting frequency for a given location $\boldsymbol{p}$ is

$$f_m(\boldsymbol{p}) = \min(f_c, f_d) = \min\left(\frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\alpha(e+e_2)}, \frac{\pi N v}{360}\right). \tag{6}$$

Therefore, the visual sensitivity model based on fovea in the spatial domain can be defined as

$$S_f(v,f,\boldsymbol{p}) = \begin{cases} \frac{CT(f,0)}{CT(f,e)} = \exp(-0.0461 f \cdot e), & f \le f_m(\boldsymbol{p}) \\ 0, & f > f_m(\boldsymbol{p}). \end{cases} \tag{7}$$

For a given image, the visual sensitivity of any point $\boldsymbol{p}$ in the spatial domain can be calculated by (7). However, most of compression methods are carried out in the wavelet domain. For the 9/7 DWT, the error detection thresholds for each subband can be represented as

$$T_{\lambda,\theta} = \frac{a 10^{k\left(\log(g_\theta f_0 2^\lambda)^2\right)}}{A_{\lambda,\theta}}. \tag{8}$$

Here, a, k, f_0, and g_θ are all constants. $A_{\lambda,\theta}$ is the amplitude of the DWT 9/7 basis function corresponding to level λ and orientation θ, and r is the display resolution. The values of a, k, f_0, g_θ, and $A_{\lambda,\theta}$ can be seen in [38].

Therefore, the error sensitivity of subband (λ, θ) can be defined as

$$S_w(\lambda,\theta) = \frac{1}{T_{\lambda,\theta}}. \tag{9}$$

Equation (9) provides an error sensitivity value for each subband in the wavelet domain. Combined with the visual sensitivity model of any point $\boldsymbol{p}$ in the spatial domain, the visual sensitivity model in the DWT domain can be calculated as

$$S(v,\boldsymbol{p}) = [S_w(\lambda,\theta)]^{\beta_1} \cdot [S_f(v,f,\boldsymbol{p})]^{\beta_2}. \tag{10}$$

$\boldsymbol{p}$ denotes the position of wavelet coefficient in subband (λ,θ). β_1 and β_2 are parameters that are used to control the magnitudes of visual sensitivity of subband and that of spatial domain based on fovea, respectively. β_1 and β_2 are estimated by our experiments. In this paper, we set $\beta_1 = 0.01$, $\beta_2 = 3$.

In Fig. 2, we can observe that $\tan(e) = u/v = d/Nv$, i.e., $d = Nv\tan(e) \approx Nve$. Here, e is in radians. In the HVS, the highest visual acuity is limited to the size of the fovea region and covers 2° of visual angle (in radians) [21]. Visual angle is the retinal eccentricity e in radians. Therefore, $d = Nv\pi/90$. This means that the fovea region is a circle with radius d.

For a fovea region, it can be regarded as the collection of multiple foveation points. Suppose that there are k foveation points $\boldsymbol{p}_1^f, \boldsymbol{p}_2^f, \ldots, \boldsymbol{p}_k^f$. For each coefficient $\boldsymbol{p}$, its visual sensitivity model $S_i(v,\boldsymbol{p}), i = 1,2,\ldots,k$ can be calculated by (10). Finally, the visual sensitivity model of it can be determined by

$$S(v,\boldsymbol{p}) = \max_{i=1\ldots,k}\left(S_i(v,\boldsymbol{p})\right). \tag{11}$$

B. Importance Weighting Mask

The purpose of the importance weight mask is to guarantee that the bits with greater contribution to the visual quality can be encoded and transmitted in priority. Based on (10) in Section II-A, the visual sensitivity is closely related to the viewing distance. However, in practical applications, the viewing distance v is not known for the encoder. Wang and Bovik [36] solved this problem by taking the probability distribution of viewing distance into consideration when calculating the weights of coefficients.

The probability distribution model of viewing distance can be represented as

$$p(v) = \frac{1}{\sqrt{2\pi}\sigma v}\exp\left(-\left(\frac{(\ln v - \mu)^2}{2\sigma^2}\right)\right), \quad v \in (0,+\infty). \tag{12}$$

Here, σ is 0.4, and μ is 1.2586. The curve of the probability distribution of viewing distance is shown in Fig. 3. Then, the importance weight of coefficient $\boldsymbol{p}_i$ is given by

$$W(\boldsymbol{p}_i) = \int_0^{+\infty} p(v)S(v,\boldsymbol{p}_i)d\,(v). \tag{13}$$

Suppose that the fovea point is the center of an image. Based on (11)–(13), the importance weighting mask is obtained, which is shown in Fig. 4.

C. HAS Method

Supposing that a transformed image is X, the size of it is $M \times N$. The scanning process of the image can be defined as a bijection b that is from a closed interval $[1,2,\ldots,M\times N]$ to the set of ordered pairs $\{(i,j) : 1 \le i \le M, 1 \le j \le N\}$,

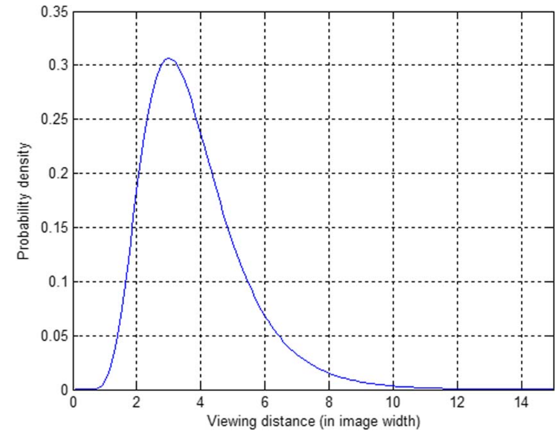


Fig. 3. Probability distribution model of viewing distance v.

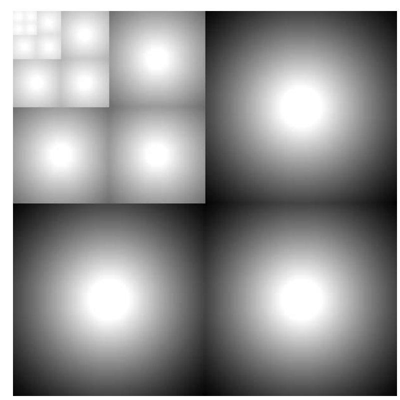


Fig. 4. Importance weighting mask.

where the latter set represents the locations in the image [39]. After scanning, the 2-D transformed image is converted to a 1-D coefficient sequence that can be represented as $[X_{b(1)}, X_{b(2)}, \ldots, X_{b(MN)}]$. The various-scanning method, in fact, is a bijection function b having different definitions. Once the definition b of a scanning method is given, then the process of compressing a 2-D image becomes a process of compressing a 1-D sequence $\tilde{X} = [X_{b(1)}, X_{b(2)}, \ldots, X_{b(MN)}]$.

For a transformed image, the coefficients that scanned first are encoded in priority. Typically, the organization of codestream is performed according to the order of scanning. This means that the codestream of those coefficients that scanned first is in front of the entire codestream. As a result, this part of the codestream will be decoded in priority and displayed if necessary. Therefore, at the same bit rate, different scanning order can lead to different quality of reconstructed image. Thus, the scanning order is very important. If those coefficients that can make a greater contribution to image reconstruction can be scanned first, then the quality of reconstructed images must be improved.

Generally, a 2-D image can be converted into a 1-D sequence by some canonical scanning strategies, such as raster scan, zigzag scan, and Morton scan. Different scanning strategies are suitable for different applications. The context-based predictive techniques, which are typically designed for lossless compression, commonly adopt the raster scan. Memon *et al.* [39]

point out that the raster scan is superior to other scanning methods, when the prediction-based techniques are used. For the transform-based coding methods, the zigzag scan can group the coefficients efficiently when the DCT transform is used. The DCT-based encoder can be regarded as the compression of a stream of 8 × 8 blocks of an image sample. For each 8 × 8 block, the DCT can concentrate most of the information of the block in the lower spatial frequencies, i.e., the top left corner of the block. After quantization, all of the 64 quantized coefficients are ordered into the zigzag sequence [48]. The reason for the zigzag scan is that it can scan the low-frequency nonzero coefficients before scanning the high-frequency coefficients. However, the zigzag scan is essentially a kind of "line" scan, i.e., instead of representing some insignificant "block," it can only represent an insignificant "line." For DWT, there are several ways to decompose an image into various subbands. Among them, octave-band decomposition is the most widely used. For the octave-band wavelet decomposition, each coefficient in a high-frequency subband has four "children" corresponding to its spatial position in the higher frequency subband. This implied that if a wavelet coefficient at a coarse scale is insignificant, then all wavelet coefficients of the same orientation in the same spatial location at a finer scale are likely to be insignificant. Therefore, for a wavelet transformed image, there could be a great many insignificant "blocks" in the wavelet domain. If these insignificant "blocks" can be organized and represented in a properly way, the coding performance will be improved. Lewis and Knowles [49] point out that the treelike data structure is an effective way to represent the coefficients of the octave decomposition. For the treelike data structure, Morton scan is the most commonly used scanning method. From the perspective of information preservation, as compared with the zigzag scan, Morton scan is a kind of "block" scanning method, which can preserve the "block" structure of the octave decomposition and thus improve the efficiency of encoding. Some classic coding algorithms, such as EZW and SPIHT, are designed with the Morton scan. However, it does not take the characteristics of wavelet subband into consideration. Moreover, the remote sensing image usually contains a great number of ground objects, which makes a lot of information still exist in high-frequency subbands after the wavelet transform, and the amount of information of these subbands may be quite different. If the fixed scanning method is adopted, it may not guarantee that the scan is performed according to the importance of the subbands. In other words, the subbands that scanned before may be less information. Therefore, the fixed scanning mode may bring a great influence on the coding performance.

Based on the aforementioned analysis, for the scanning of a transformed image, the "block" scanning method is more appropriate. A question then arises: for a remote sensing image, what is a proper "block" scanning method? In addition, for the problem that the amount of information of high-frequency subbands may be quite different, how to design an effective scanning order among subbands is another question that needs to be considered.

In this paper, we present a novel HAS scheme for the compression of remote sensing images. The process of the HAS is presented in Algorithm 1.

Algorithm 1 The process of HVS-based, adaptive scanning method.

Input: A wavelet transformed image X, the level of decomposition is J.

- Calculate the retina-based visual sensitivity model, and then weight the image X by the generated importance weighting mask W. The weighted image can be represented as

$$X_w = X \cdot W.$$

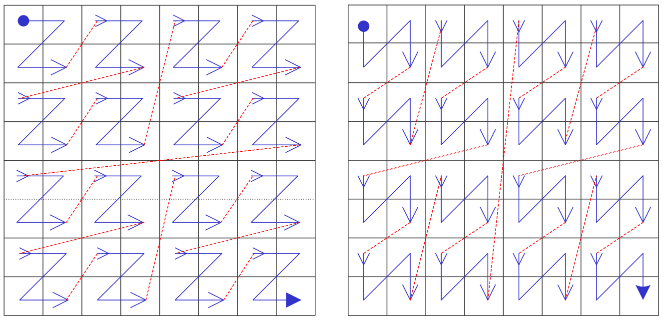
- Calculate the energy of each weighted subband [43]. The energy can be represented it as $E_{m,d}$. m-Scale $(m = 1, 2, \ldots, J)$, d-Direction $(d = 1, 2, 3, 4$. "1" represents the lowest frequency subband. "2" represents the horizontal direction, "3" represents the diagonal direction, "4" represents the vertical direction, respectively)

$$E_{m,d} = \sum_{i,j}^{R,C} c_w(i,j)^2$$

  R and C are the number of row and column of the current subband, respectively; $c_w(i,j)$ represents the coefficient of the current weighted subband that located in (i,j).
- Determine the scanning order among all the weighed subbands according to their energy.
- Determine the scanning order of each weighed subband.
  For each weighted subband $X_w(m,d)$
  — If $m = 1$ or $m = 2$, then the "horizontal z-scan" is adopted.
  — Else if $m = 4$, then the "vertical z-scan" is exploited.
  — Else if $m = 3$, then the scan method depends on the horizontal subband and vertical subband of this level.
    1) If $E_{m,2} \geq E_{m,4}$, the "horizontal z-scan" is performed to this subband.
    2) If $E_{m,2} < E_{m,4}$, the "vertical z-scan" scan is performed to this subband.
  End

Output: 1-D coefficients sequence that generated by scanning the 2-D transformed image.

From Algorithm 1, it has been demonstrated that the proposed HAS scheme is mainly performed by three steps. First, for the transformed image, all the wavelet coefficients are weighted by the importance weighting mask. Second, the energy of each weighted subband is calculated, and the scanning order among subbands is determined according to their energy in descending order. The purpose of doing so is to organize the codestream, according to the importance of weighted subbands. Finally, for the scanning within a weighted subband, we take the characteristics of subbands into consideration. For the horizontal subbands, they reflect the information of an image in the horizontal direction; hence, we adopt the "horizontal z-scan"



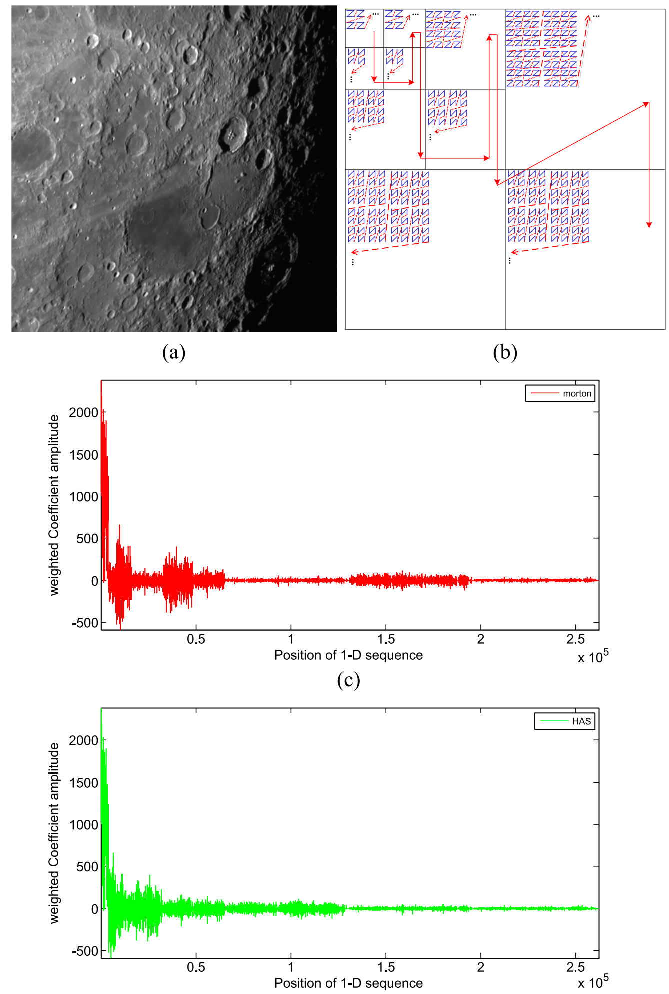
(a) (b)

Fig. 5. "Horizontal z-scan" method and the "vertical z-scan" method. (a) "Horizontal z-scan" method used to the subbands with more horizontal information. (b) "Vertical z-scan" method that used to the subbands with more vertical information.

as the scanning method. Similarly, for those vertical subbands, they represent the vertical information of an image; if the scan can be performed along the vertical direction, coding performance should be improved. The "vertical z-scan" method is designed for the vertical subbands in this paper. The "horizontal z-scan" method and the "vertical z-scan" method are shown in Fig. 5(a) and (b), respectively. In addition, for the diagonal subbands, the scanning method depends on the image itself, i.e., if the horizontal information of an image is much more than vertical information in the current wavelet level, then the "horizontal z-scan" method is adopted. Otherwise, the "vertical z-scan" method is conducted.

Now, we give an example of the HAS method. The original image is shown in Fig. 6(a). For the given image, the visual weighting mask is generated by utilizing the visual sensitivity model and the probability distribution of viewing distance. Then, each wavelet coefficient is weighted based on the visual weighting mask. Suppose that the level of wavelet decomposition is 3. Based on Algorithm 1, the energy of each weighted subband is calculated, and the results are tabulated in Table I.

According to Table I, the scanning order among weighted subbands is determined, i.e., LL_3, LH_3, HH_3, HL_3, LH_2, HH_2, HL_2, LH_1, HL_1, and HH_1. Then, for each subband, the scanning method of it depends on the characteristics of the subband. The "horizontal z-scan" method is used for the LL_3, HL_3, HL_2, and HL_1. The "vertical z-scan" method is exploited for the LH_3, LH_2, and LH_1. In Table I, it can be observed that the vertical information is much more than the horizontal information for each level of the transformed image. Therefore, for all the diagonal subbands, i.e., HH_3, HH_2, and HH_1, the "vertical z-scan" method is adopted. The whole process of scanning is shown in Fig. 6(b). Finally, the 1-D coefficient sequence is obtained. Fig. 6(c) and (d) shows the 1-D coefficient sequence generated by Morton scan and the proposed HAS method, respectively. In Fig. 6, we can observe that, the proposed HAS method can scan these weighted subbands adaptively, which can place those important coefficients in the front of the 1-D sequence, and preserve the texture information as much as possible. This scanning way will help to improve the visual quality of reconstructed images on the basis of visual weighting.



(a) (b) (c) (d)

Fig. 6. Original image and its process of scanning. (a) Lunar (8 bit, 512 × 512). (b) Process of HAS of the weighted transformed image of (a). (c) One-dimensional coefficient sequence obtained by Morton scan. (d) One-dimensional coefficient sequence obtained by the proposed HAS method.

TABLE I
ENERGY OF EACH SUBBAND OF THE TRANSFORMED IMAGE($\times 10^9$)

Subband	Energy	Subband	Energy	Subband	Energy	Subband	Energy
LL_3	4.6745	HL_3	0.0075	HH_3	0.0140	LH_3	0.0459
---	---	HL_2	0.0058	HH_2	0.0072	LH_2	0.0406
---	---	HL_1	0.0016	HH_1	0.0009	LH_1	0.0124

D. Overhead of Bits

In Section II-A and B, it is worth noting that the importance weighting masking is independent of the content of an image. Thus, the mask is not needed to be transmitted to the receiver. Only some side information, including the fovea point and the width of the image, is needed to be sent for decoding correctly. In this paper, four integers are used to record the position of the fovea point (i.e., two integers for its horizontal ordinate and the other two integers for its vertical ordinate), and two integers are used to record the width of the image. Thus, the side information of visual weighting mask is only six integers.

In addition, for the adaptive scanning method, the side information, including the scanning order among the subbands and the scanning method of those diagonal subbands, is also needed to be sent to the receiver. Supposing that the level of wavelet decomposition is J, only $3J+1$ integers are needed to represent the scanning order among the subbands. In addition, J integers are needed to represent the scanning methods of those diagonal subbands. Therefore, the cost of side information of the adaptive scanning method is $(3J+1)+J=4J+1$. As a result, the total side information at the encoder is $6+(4J+1)=4J+7$.

Now, we take the image in Fig. 6(a) as an example. The size of the "lunar" is 512×512, and the level of decomposition J is 3. Supposing that the compression ratio is c_r, then the percentage of overhead bits is $(4 \times 3+7)/(512 \times 512)/c_r \approx 0.0000725 \cdot c_r$. This means that the percentage of overhead bits is proportional to the compression ratio and very small. Moreover, if the size of the image is larger, then the proportion of overhead bits would be smaller. Furthermore, if the entropy coding is used to encode these overhead bits, the cost of overhead would be reduced further. Based on the aforementioned analysis, the overhead bits of the proposed HAS-based method are extremely small and nearly can be negligible.

III. Binary Tree Coding

Most of the embedded coding methods are based on quadtree decomposition, such as EZW [3], SPIHT [4], and SPECK [5]. However, Shaffer *et al.* [40] pointed out that the coding method based on binary tree decomposition is more efficient and simpler than those based on quadtree decomposition. The state-of-the-art compression approach based on binary tree is proposed in [35], which developed a new method called BTCA. The BTCA is extremely suitable for the compression of remote sensing images because it can preserve more details. In this paper, the BTCA is adopted as the embedded codec.

Function code $=\mathrm{BTC}(\Gamma, i, T_k)$
Input: Γ represents a binary tree, and i is the index of a node of the binary tree. T_k represents the threshold. $T_0=2^{\lfloor\log_2 \Gamma(1)\rfloor}$, and $T_k=T_0/2^k$.

- If $\Gamma(i)$ has been coded with significant in larger threshold, i.e., $\Gamma(i) \geq T_{k-1}$, then
  If $\Gamma(i)$ is not in the bottom level of the binary tree, code the two children of $\Gamma(i)$, else the sign of $\Gamma(i)$ is coded.
- If $\Gamma(i)$ has a significant parent, and the brother of $\Gamma(i)$ is insignificant, then
  If $\Gamma(i)$ is not in the bottom level of the binary tree, code the two children of $\Gamma(i)$, else the sign of $\Gamma(i)$ is coded.
- If $\Gamma(i) \geq T_k$
  If $\Gamma(i)$ is not in the bottom level of the binary tree, code the two children of $\Gamma(i)$, and add a "1" before the codestream. Else the sign of $\Gamma(i)$ is coded, and add a "1" before the codestream.
- Else
  "0" is output.

Output: The codestream of the subtree whose root is the node $\Gamma(i)$.

Function code $=\mathrm{BTCA}(T_k)$
Input: T_k is the threshold.
Initialization: $d=D$.
While $(d>1)$
{
- For $i=\sum_{j=0}^{d-1} 2^j+1\ to\ \sum_{j=0}^{d} 2^j$
- Let $ct=\{\ \}$. If $\Gamma(i) \geq T_{k-1}$
  If $\Gamma(i)$ is on the left of its brother, then $ct=\mathrm{BTC}(\Gamma, i+1, T_k)$;
  Else
  $ct=\mathrm{BTC}(\Gamma, i-1, T_k)$.
- code $=\{\text{code}, ct\}$.
- $d=d-1$.

Output: The codestream of the bit plane for the given threshold T_k.
}

The process of the binary tree coding (BTC) can be described by the function BTC. Supposing that D represents the bottom level of the tree, then the process of BTCA can be described by the function BTCA.

IV. Quality Evaluation Index

In Section I, we have pointed out that the purpose of the proposed HAS-based compression method is to meet the growing demand for browsing remote sensing images online. Therefore, the proposed compression method should be evaluated by some indexes related to human eye. In this paper, Foveation Wavelet domain Quality Index (FWQI) [36], [45], visual signal-to-noise ratio (VSNR) [41], [50], [51], and MultiScale Structural Similarity Index Measure (MS-SSIM) [42], [52], [53] are chosen as the evaluation indexes.

A. FWQI

In [43], Wang and Bovik proposed an image quality index by modeling the image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Then, they adapted the index into the DWT domain and defined the FWQI [36] as

$$\mathrm{FWQI}=\frac{\sum_{n=1}^{M} S(v, \boldsymbol{x}_n) \cdot |c(\boldsymbol{x}_n)|\, Q(\boldsymbol{x}_n)}{\sum_{n=1}^{M} S(v, \boldsymbol{x}_n) \cdot |c(\boldsymbol{x}_n)|}. \tag{14}$$

Here, M is the number of the wavelet coefficients. $c(\boldsymbol{x}_n)$ represents the wavelet coefficients at location x_n. $Q(\boldsymbol{x}_n)$ represents the quality value at location $\boldsymbol{x}_n$ in the quality index map. Since $S(v, \boldsymbol{x}_n)$ varies with v, thus, the FWQI of a test image is a function of v, instead of a single value.

B. VSNR

Chandler and Hemami [41] presented an efficient metric, i.e., the VSNR metric, for quantifying the visual fidelity of an image based on near-threshold and suprathreshold properties of human vision. The VSNR is generally competitive with

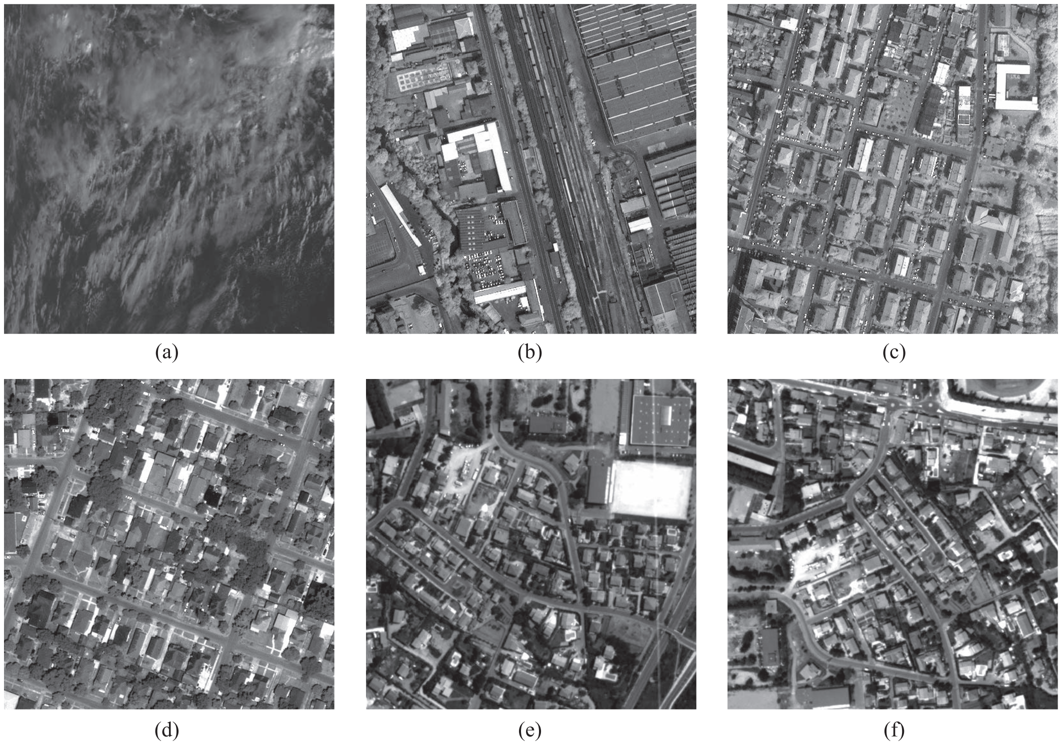


Fig. 7. Parts of the remote sensing images used in the experiment. (a) ocean_2kb1. (b) pavia1. (c) pavia2. (d) houston. (e) pleiades_portdebouc_pan1. (f) pleiades_portdebouc_pan2.

other metrics of visual fidelity. The VSNR, in decibels, can be calculated as follows:

$$\mathrm{VSNR} = 20\log_{10}\left(\frac{C(f)}{\alpha d_{\mathrm{pc}} + (1-\alpha)\frac{d_{\mathrm{gp}}}{\sqrt{2}}}\right) \tag{15}$$

where $C(f)$ denotes the RMS contrast of the original image f, d_{pc} denotes the measure of the perceived contrast of the distortions, and d_{gp} presents the measure of the extent to which global precedence has been disrupted. α is set to 0.04 for providing a reasonable fit to subjective rating data.

C. MS-SSIM

The MS-SSIM is a multiscale structural similarity method, which incorporates the variations of viewing conditions and is more flexible than Structural Similarity Index Measure (SSIM). Thus, we adopt the MS-SSIM as an evaluation index in this paper. That is

$$\mathrm{SSIM}(x,y) = [l_M(x,y)]^{\alpha_M} \cdot \prod_{j=1}^{M}\left[c_j(x,y)^{\beta_j} s_j(x,y)^{\gamma_j}\right]. \tag{16}$$

Here, $l(x,y)$, $c(x,y)$, and $s(x,y)$ represent the luminance, contrast, and structure comparison, respectively.

V. Experiments and Results

Here, some experiments are implemented to verify the performance of the proposed HAS-based compression method for remote sensing images. The results of the proposed method are compared with that of other scan-based compression methods at different bit rates.

A. Experimental Data Set

In order to prove the efficiency of the proposed HAS-based compression method, several remote sensing images with different bit depths are chosen in the experiments. Most of them are with high spatial resolution.

Some test images are from the CCSDS reference test image set [44], including "lunar," "coastal-b1," "ocean-2kb1," and "pleiades_portdebouc_pan." We crop the upper left of these images with the size of 512 $\times$ 512, for comparison under the same condition. Moreover, in order to fully verify the effectiveness of the proposed method, two other remote sensing images are chosen. The image "pavia" comes from the image acquired by the QuickBird sensor over Pavia, Northern Italy, whose resolution is 0.6 m, and "houston" comes from the image acquired by the WorldView-2 sensor over Houston, TX, USA, in 2013, whose resolution is 0.5 m. The sizes of test image "pavia" and "houston" are all 512 $\times$ 512. In the test image set, the bit depth of "lunar" and "coastal-b1" is 8 bits, the bit depth of "ocean_2kb1" is 10 bits, the bit depth of "pavia" and "houston" is 11 bits, and the bit depth of "pleiades_portdebouc_pan" is 12 bits. Parts of the test images are shown in Fig. 7.

B. Comparison of Visual Quality

In order to compare the visual quality of reconstructed images obtained by the proposed HAS-based compression method and that obtained by other scan-based compression methods,

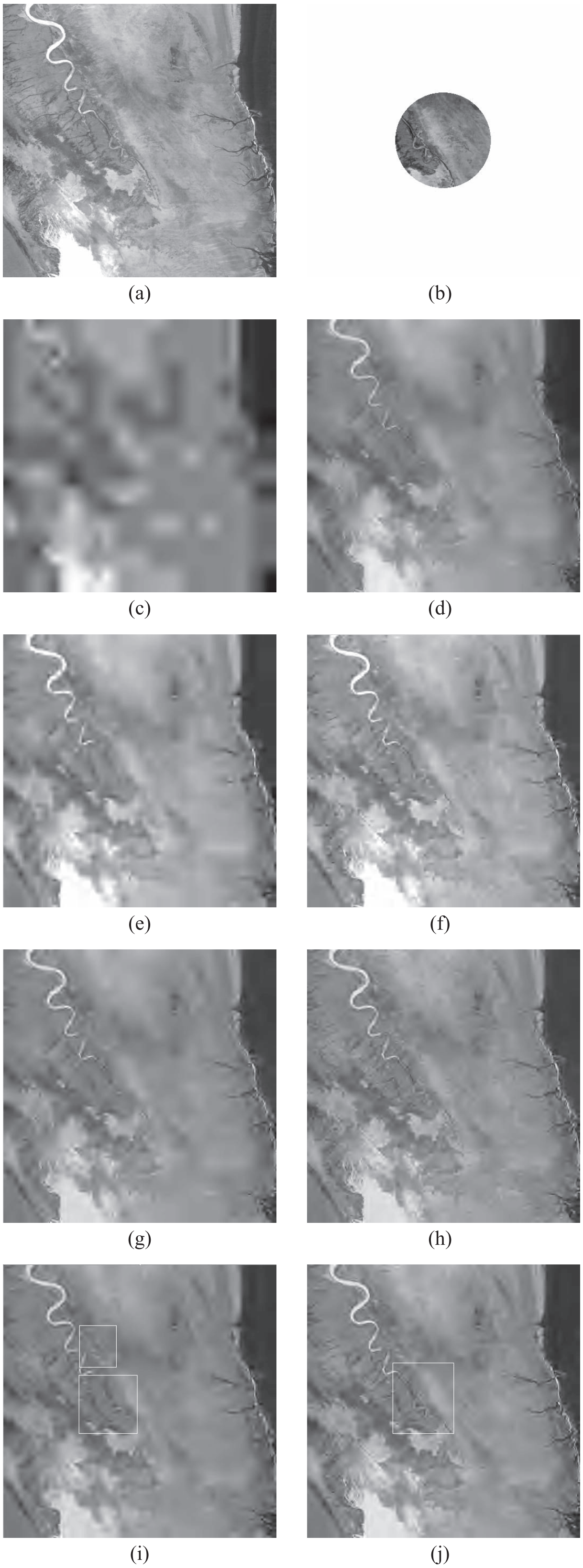


Fig. 8. Comparison of visual quality of the proposed HAS-based compression method and other scan-based compression methods with the "coastal-b1" image. (a) Original image. (b) Fovea region of (a) when $v = 5$. (c) and (d) Reconstucted images obtained by SPIHT at 0.0313 and 0.0625 bpp, respectively. (e) and (f) Reconstucted images obtained by JPEG2000 at 0.0313 and 0.0625 bpp, respectively. (g) and (h) Reconstucted images obtained by BTCA at 0.0313 and 0.0625 bpp, respectively. (i) and (j) Reconstucted images obtained by the proposed method at 0.0313 and 0.0625 bpp, respectively.

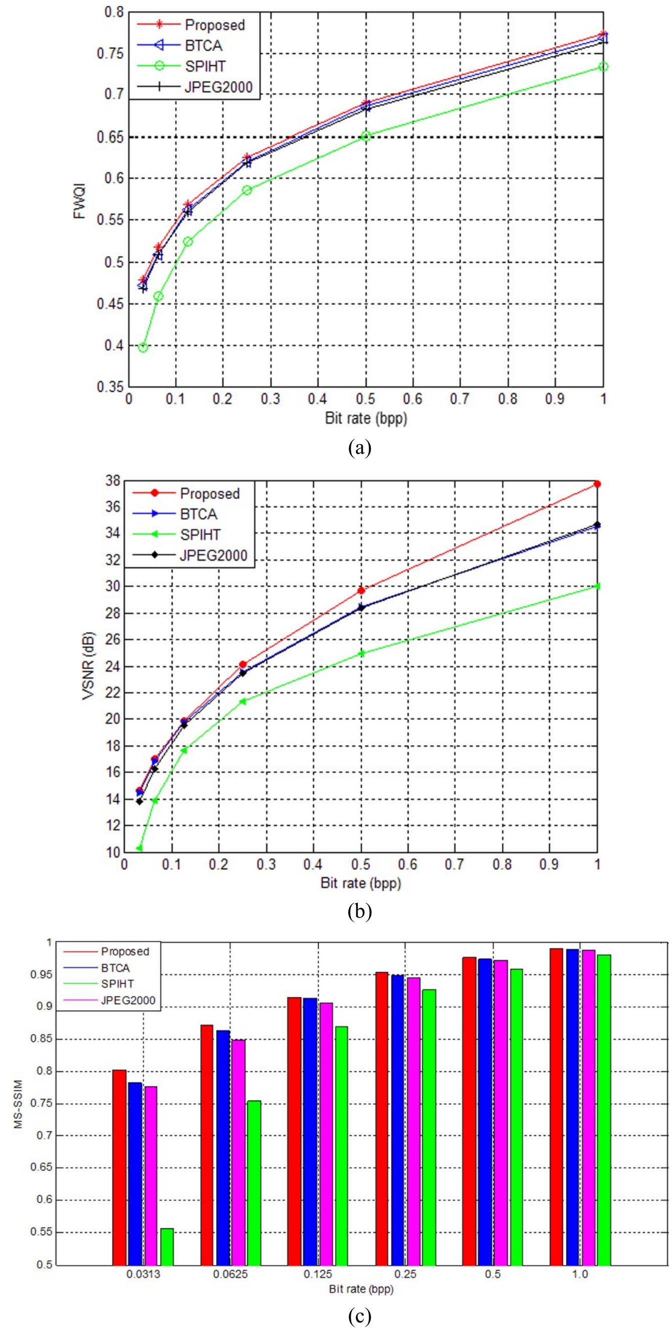


Fig. 9. Result of some quality evaluation indexes at different bit rates. (a) Result of the FWQI. (b) Result of the VSNR. (c) Result of the MS-SSIM.

we choose the "coastal-b1" as test image. Supposing that the fovea point is the center of an image and the viewing distance v is 5, the level of wavelet decomposition is 5. The image is compressed by SPIHT, JPEG2000, BTCA, and the proposed HAS-based compression method, respectively. The comparison results of the visual quality of reconstructed images at different bit rates are shown in Fig. 8.

In Fig. 8, we can observe that the visual quality of reconstructed images obtained by SPIHT is the worst. For JPEG2000, some sophisticated components, such as context modeling and rate–distortion optimization, particularly the HVS-based CSF model, all can help it to obtain a good coding performance. Therefore, the visual quality of reconstructed images obtained

TABLE II
FWQI FOR THE PROPOSED COMPRESSION METHOD AND OTHER SCAN-BASED COMPRESSION METHODS

Image	0.0313bpp				0.0625bpp				0.125bpp			
	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed
lunar	0.260	0.323	0.338	0.341	0.325	0.373	0.387	0.389	0.399	0.452	0.451	0.455
coastal-bl	0.397	0.468	0.472	0.478	0.459	0.508	0.508	0.518	0.524	0.560	0.563	0.569
ocean_2kb1	0.312	0.376	0.393	0.398	0.389	0.440	0.447	0.452	0.466	0.510	0.517	0.521
pavia1	0.338	0.398	0.411	0.418	0.400	0.438	0.444	0.447	0.454	0.494	0.495	0.498
pavia2	0.322	0.397	0.405	0.408	0.391	0.450	0.451	0.455	0.460	0.510	0.513	0.517
houston	0.312	0.381	0.389	0.391	0.384	0.433	0.436	0.438	0.452	0.500	0.496	0.502
pleiades1	0.326	0.395	0.417	0.419	0.417	0.464	0.475	0.476	0.504	0.549	0.549	0.555
Pleiades2	0.301	0.378	0.397	0.399	0.389	0.451	0.457	0.459	0.485	0.532	0.537	0.545
Average	0.321	0.390	0.403	**0.407**	0.394	0.445	0.451	**0.454**	0.468	0.513	0.515	**0.520**

TABLE III
FWQI FOR THE PROPOSED COMPRESSION METHOD AND OTHER SCAN-BASED COMPRESSION METHODS

Image	0.25bpp				0.5bpp				1bpp			
	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed
lunar	0.495	0.538	0.538	0.539	0.592	0.628	0.631	0.633	0.709	0.739	0.746	0.748
coastal-bl	0.586	0.619	0.621	0.626	0.651	0.682	0.686	0.690	0.733	0.763	0.768	0.773
ocean_2kb1	0.554	0.594	0.595	0.598	0.641	0.684	0.676	0.688	0.744	0.776	0.776	0.784
pavia1	0.521	0.557	0.555	0.561	0.597	0.635	0.627	0.638	0.686	0.727	0.717	0.722
pavia2	0.540	0.584	0.586	0.589	0.624	0.657	0.656	0.659	0.709	0.746	0.737	0.739
houston	0.526	0.561	0.562	0.569	0.609	0.650	0.642	0.646	0.697	0.735	0.727	0.731
pleiades1	0.599	0.640	0.641	0.649	0.693	0.735	0.732	0.737	0.785	0.811	0.815	0.818
Pleiades2	0.586	0.629	0.631	0.635	0.686	0.732	0.721	0.730	0.776	0.804	0.812	0.816
Average	0.551	0.590	0.591	**0.596**	0.637	0.675	0.671	**0.678**	0.730	0.763	0.762	**0.766**

by JPEG2000 is better than that of SPIHT. However, as mentioned in Section I, The existing Part 1 JPEG2000-compliant visual progressive weighting scheme only uses a single weight for an entire coding block, whose size is 64 × 64 or 32 × 32. For each code block, it corresponds to a large portion of the relevant image. Theoretically, the visual sensitivities of these coefficients in this portion are different; each coefficient should be weighted by a weight that related to its visual sensitivity. Therefore, a main limitation of the JPEG2000 visual weighting mask is that the weight is not very fine enough. It will affect the visual quality of reconstructed images. For BTCA, the binary tree coding is more efficient, which traverses the binary tree level by level from the bottom to the top, and also can provide a good visual effect. The visual quality of reconstructed images obtained by the proposed HAS-based compression method is the best. The reason is that the proposed method adopts the retina-based visual sensitivity model, which provides a weight for each coefficient of the image based on the HVS. It helps to guarantee that the bits with greater contribution to human vision are encoded and transmitted in priority, particularly those bits in the fovea region. In addition, both the adaptive scanning and encoding processes of the proposed compression method can help to preserve more texture information. As a result, the visual quality of reconstructed images must be improved. It can be observed that, in Fig. 8(i) and (j), for a given bit rate, as compared with other methods, the details in the rectangle boxes that belong to the fovea region and its surrounding area of the reconstructed image obtained by the proposed method are more abundant and clearer. This proves that the proposed compression method is more consistent with the characteristics of human vision.

In order to verify the effectiveness of the proposed method further, for the image in Fig. 8(a), under the same condition, more experiments are performed. The result of FWQI, VSNR, and MS-SSIM of all these methods at bit rates from 0.0313 to 1 bpp are shown in Fig. 9(a)–(c), respectively. It has been demonstrated that, from the perspective of objective evaluation, the visual quality of reconstructed images obtained by the proposed HAS-based compression method is still superior to that obtained by other scan-based compression methods at all bit rates.

C. Performance Comparison of the Proposed HAS-Based Compression Method With Other Scan-Based Methods

Here, more experiments are conducted to verify the proposed method. In these experiments, all test images mentioned in Section V-A are used. They are decomposed by five-level 9/7-tap biorthogonal wavelet filters. In Fig. 3, we can observe that the maximum possible viewing distance v is 3. Therefore, for these experiments, v is set to 3. The result of FWQI, VSNR, and MS-SSIM of all compression methods at different bit rates are tabulated in Tables II–VII, respectively.

FWQI is a quality index relevant to the fovea point, and it is a function of viewing distance. It can evaluate the proposed compression method effectively. In Tables II and III, we can observe that, when the bit rate is less than 0.5 bpp, for all the test images, the FWQIs of reconstructed images obtained by the proposed HAS-based compression method are better than that of all other compression methods. The reason is that, for the proposed method, the retina-based visual weighting masking is adopted, and the adaptive scanning strategy takes the content of image and the characteristics of subband into account, simultaneously. As a result, comparing with other compression methods, more visual important information can be preserved, which is beneficial to improve the visual effect of

TABLE IV
VSNR (dB) FOR THE PROPOSED COMPRESSION METHOD AND OTHER SCAN-BASED COMPRESSION METHODS

Image	0.0313bpp				0.0625bpp				0.125bpp			
	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed
lunar	10.561	11.935	12.649	12.654	12.897	14.021	14.690	14.752	15.747	17.401	17.662	17.820
coastal-bl	10.246	13.846	14.427	14.593	13.909	16.263	16.853	17.016	17.685	19.510	19.784	19.887
ocean_2kb1	10.555	12.721	13.527	13.570	13.870	15.522	15.954	16.000	16.883	18.622	19.134	19.313
pavia1	12.402	13.944	14.245	14.327	14.324	15.508	16.083	16.124	16.763	18.623	18.611	18.733
pavia2	11.092	12.272	12.423	12.768	12.796	13.847	14.008	14.459	14.787	16.470	16.139	16.548
houston	11.677	13.274	13.855	13.960	14.360	15.692	15.936	16.113	17.051	18.938	18.802	19.107
pleiades1	10.515	12.381	13.008	13.055	13.647	14.957	15.492	15.562	17.069	19.029	19.317	19.371
Pleiades2	8.586	10.174	10.870	10.893	11.517	12.992	13.366	13.429	14.858	16.183	16.889	16.958
Average	10.704	12.568	13.126	**13.228**	13.415	14.850	15.298	**15.432**	16.355	18.097	18.292	**18.467**

TABLE V
VSNR (dB) FOR THE PROPOSED COMPRESSION METHOD AND OTHER SCAN-BASED COMPRESSION METHODS

Image	0.25bpp				0.5bpp				1bpp			
	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed
lunar	19.554	22.061	22.124	22.186	24.291	26.670	27.722	28.035	30.893	35.250	37.556	38.719
coastal-bl	21.342	23.441	23.519	24.132	24.914	28.361	28.442	29.718	30.021	34.700	34.555	37.714
ocean_2kb1	20.787	23.612	23.388	24.022	25.562	30.417	29.069	31.056	31.285	39.228	36.939	41.086
pavia1	20.017	22.027	22.079	22.735	23.911	27.563	26.827	28.138	29.193	34.544	33.743	35.905
pavia2	17.673	19.981	19.438	19.839	21.568	25.443	24.385	24.954	27.198	32.797	31.579	33.460
houston	20.445	23.076	22.532	23.212	25.080	29.442	28.581	29.384	31.079	39.318	36.534	38.514
pleiades1	21.294	24.593	24.328	24.512	27.027	33.267	33.002	33.346	36.217	44.110	42.613	43.880
Pleiades2	19.165	22.137	22.240	22.727	25.117	30.955	29.562	30.575	31.561	40.646	39.829	41.357
Average	20.035	22.616	22.456	**22.921**	24.684	29.015	28.449	**29.401**	30.931	37.574	36.669	**38.829**

TABLE VI
MS-SSIM FOR THE PROPOSED COMPRESSION METHOD AND OTHER SCAN-BASED COMPRESSION METHODS

Image	0.0313bpp				0.0625bpp				0.125bpp			
	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed
lunar	0.485	0.670	0.721	0.730	0.689	0.785	0.803	0.808	0.812	0.872	0.878	0.887
coastal-bl	0.556	0.777	0.782	0.803	0.754	0.848	0.863	0.871	0.869	0.907	0.913	0.914
ocean_2kb1	0.440	0.683	0.735	0.746	0.699	0.808	0.822	0.831	0.829	0.888	0.893	0.901
pavia1	0.212	0.477	0.483	0.515	0.436	0.612	0.647	0.679	0.652	0.769	0.769	0.778
pavia2	0.214	0.516	0.518	0.561	0.463	0.652	0.673	0.688	0.664	0.791	0.771	0.797
houston	0.294	0.541	0.594	0.570	0.548	0.693	0.712	0.701	0.722	0.823	0.804	0.830
pleiades1	0.322	0.610	0.663	0.649	0.644	0.757	0.775	0.793	0.813	0.877	0.882	0.888
Pleiades2	0.322	0.614	0.665	0.677	0.650	0.770	0.788	0.803	0.816	0.871	0.890	0.894
Average	0.356	0.611	0.645	**0.656**	0.610	0.741	0.760	**0.772**	0.772	0.850	0.850	**0.861**

TABLE VII
MS-SSIM FOR THE PROPOSED COMPRESSION METHOD AND OTHER SCAN-BASED COMPRESSION METHODS

Image	0.25bpp				0.5bpp				1bpp			
	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed	SPIHT	J2K	BTCA	Proposed
lunar	0.902	0.931	0.941	0.947	0.954	0.965	0.972	0.975	0.982	0.988	0.991	0.992
coastal-bl	0.926	0.945	0.948	0.954	0.959	0.972	0.974	0.977	0.981	0.987	0.989	0.990
ocean_2kb1	0.910	0.942	0.943	0.949	0.956	0.974	0.975	0.977	0.982	0.990	0.991	0.991
pavia1	0.805	0.865	0.867	0.883	0.899	0.938	0.938	0.942	0.953	0.972	0.973	0.974
pavia2	0.809	0.882	0.875	0.887	0.904	0.944	0.944	0.946	0.959	0.977	0.977	0.978
houston	0.839	0.896	0.893	0.902	0.918	0.951	0.951	0.955	0.963	0.979	0.978	0.982
pleiades1	0.907	0.943	0.945	0.948	0.960	0.977	0.979	0.982	0.986	0.991	0.992	0.993
Pleiades2	0.913	0.948	0.951	0.952	0.964	0.980	0.981	0.982	0.986	0.993	0.994	0.994
Average	0.876	0.919	0.920	**0.928**	0.939	0.963	0.964	**0.967**	0.974	0.985	0.986	**0.987**

reconstructed images. On the other hand, with the increasing bit rate, some sophisticated algorithms, such as JPEG2000, sometimes provide a better result. For example, Table III shows that, for the test image "houston," when the bit rate is 0.5 bpp, the FWQI of the proposed method is 0.646, and that of the JPEG2000 is 0.650. The reason is that, when the bit rate is high, some sophisticated components of JPEG2000, such as context modeling and rate–distortion optimization, all can help it to obtain a better performance. However, the slight performance improvement is at the cost of high complexity. Nevertheless, for most of the test images, the proposed HAS-based method is still the best compression method at high bit rates. Typically, different images are of different contents, including the complexity and texture. From the perspective of evaluating an algorithm, instead of evaluating it by one image, a better way is to use many test images and evaluate the average result. According to Tables II and III, for all given bit rates, the average FWQIs of the proposed method are the highest. This means that the proposed HAS-based compression method is more effective compared with other compression methods.

Tables IV and V list the VSNR results of the proposed method and the other three methods. Based on the aforementioned analysis, the proposed compression method is still superior to other methods. As an error summation method, MS-SSIM can provide a whole approximation to the perceived image quality from the perspective of structural similarity. According to Tables VI and VII, we can observe that the MS-SSIM results of the proposed HAS-based compression method are still better than those of other scan-based methods.

According to the results listed in Tables II–VII, we can observe that, for all given bit rates, the average results of FWQI, VSNR, and MS-SSIM obtained by the proposed compression method are all the best. This illustrates that, as compared with other scan-based compression methods, the proposed HAS-based compression method can provide better visual quality of reconstructed images.

VI. Conclusion and Discussion

In this paper, we have presented a HAS scheme for the compression of remote sensing images. Different scanning orders among subbands and within a subband are designed, respectively. The human visual sensitivity, the content of image, and the characteristics of remote sensing images are considered in the proposed compression algorithm, simultaneously. The overhead bits of the proposed HAS-based compression method are extremely small and almost can be negligible. The proposed compression method can achieve the progressive image transmission because of its binary tree strategy. Experimental results show that, as compared with other scan-based methods, the proposed compression method can provide better visual quality of reconstructed images in each phase. In addition, the proposed method does not adopt any sophisticated components such as context modeling or rate–distortion optimization, even entropy coding. Therefore, it is of low complexity, which means that it can be implemented in hardware or software easily. The proposed HAS-based compression method is very suitable for the vision-related applications for remote sensing images, such as the remote sensing map online browsing, and has a very broad application prospect.

The determination of fovea point is a problem worth studying. In this paper, the fovea point is assumed to be the center of an image, which is under the premise that the user is not particularly interested in a certain area. In practical applications, the fovea point may be not always the center of an image. Theoretically, there are two ways to determine the fovea region. One is the completely automatic method, which can locate the region of interest automatically. However, it is very difficult to implement. The reason is that, for remote sensing images, it usually contains the information of terrain or landform; unlike some common applications such as face recognition, the target of remote sensing images is hard to determine. Another more feasible approach is to ask users to indicate the point or region of interest and then feedback the information to the encoder. However, this requires that the compression system is a real-time system. How to provide a good visual quality, while speeding up the process of real-time compression, is an issue worthy studying. We will do some research on it in the future.

References

[1] J. M. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] Y. H. Liu *et al.*, "Applying GPU and POSIX thread technologies in massive remote sensing image data processing," in *Proc. Geoinformat. Conf.*, Beijing, China, 2011, pp. 1–6.

[3] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[4] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, Jun. 1996.

[5] W. A. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient low complexity image coding with a set-partitioning embedded block coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 11, pp. 1219–1235, Nov. 2004.

[6] *JPEG2000 Image Coding System*, ISO/IEC Std. 15 444-1, 2000.

[7] S. Patel and S. Srinivasan, "Modified embedded zerotree wavelet algorithm for fast implementation of wavelet image codec," *Electron. Lett.*, vol. 36, no. 20, pp. 1713–1714, Sep. 2000.

[8] V. N. Ramaswamy, K. R. Namuduri, and N. Ranganathan, "Context-based lossless image coding using EZW framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 554–559, Apr. 2001.

[9] S. R. Chang and L. Carin, "A modified SPIHT algorithm for image coding with a joint MSE and classification distortion measure," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 713–725, Mar. 2006.

[10] Z. J. Fang, N. X. Xiong, L. T. Yang, X. M. Sun, and Y. Yang, "Interpolation-based direction-adaptive lifting DWT and modified SPIHT for image compression in multimedia communications," *IEEE Syst. J.*, vol. 5, no. 4, pp. 584–593, Dec. 2011.

[11] Y. Jin and H. J. Lee, "A block-based pass-parallel SPIHT algorithm," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 7, pp. 1064–1075, Jul. 2012

[12] Z. Y. Wu, A. Bilgin, and M. W. Marcellin, "Joint source/channel coding for image transmission with JPEG2000 over memoryless channels," *IEEE Trans. Image Process.*, vol. 14, no. 8, pp. 1020–1032, Aug. 2005.

[13] J. Y. Yang, Y. Wang, W. L. Xu, and Q. H. Dai, "Image coding using dual-tree discrete wavelet transform," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1555–1569, Sep. 2008.

[14] L. Z. Wang, Y. Ma, A. Y. Zomaya, R. Ranjan, and D. Chen, "A parallel file system with application-aware data layout policies for massive remote sensing image processing in digital earth," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1497–1508, Jun. 2015.

[15] A. Beghdadi, M. C. Larabi, A. Bouzerdoum, and K. M. Lftekharuddin, "A survey of perceptual image processing methods," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 811–831, Sep. 2013.

[16] B. Macq and H. Q. Shi, "Perceptually weighted vector quantization in the DCT domain," *Electron. Lett.*, vol. 29, no. 15, pp. 1382–1384, Jul. 1993.

[17] I. Hontsch and L. J. Karam, "Locally adaptive perceptual image coding," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1472–1483, Sep. 2000.

[18] I. Hontsch and L. J. Karam, "Adaptive image coding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 213–222, Mar. 2002.

[19] M. G. Albanesi and F. Guerrini, "An HVS-based adaptive coder for perceptually lossy image compression," *Pattern Recog.*, vol. 36, no. 4, pp. 997–1007, Apr. 2003.

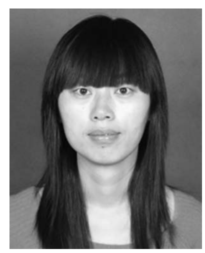
[20] M. J. Nadenau, J. Reichel, and M. Kunt, "Wavelet-based color image compression: Exploring the contrast sensitivity function," *IEEE Trans. Image Process.*, vol. 12, no. 1, pp. 58–70, Jan. 2003.

[21] Z. Liu, L. J. Karam, and A. B. Watson, "JPEG2000 encoding with perceptual distortion control," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1763–1778, Jul. 2006.

[22] G. Sreelekha and P. S. Sathidevi, "An HVS based adaptive quantization scheme for the compression of color images," *Digital Signal Process.*, vol. 20. no. 4. pp. 1129–1149, Jul. 2010.

[23] D. Wu *et al.*, "Perceptually lossless medical image coding," *IEEE Trans. Medical Image*, vol. 25, no. 3, pp. 335–344, Mar. 2006.

[24] X. H. Zhang, W. S. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *J. Vis. Commun. Image R*, vol. 19, no. 1, pp. 30–41, Jan. 2008.
[25] D. M. Tan, C. S. Tan, and H. R. Wu, "Perceptual color image coding with JPEG2000," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 374–383, Feb. 2010.
[26] H. Oh, A. Bilgin, and M. W. Marcellin, "Visually lossless encoding for JPEG2000," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 189–201, Jan. 2013.
[27] Y. Niu, X. L. Wu, G. M. Shi, and X. T. Wang, "Edge-based perceptual image coding," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1899–1910, Apr. 2012.
[28] A. L. N, T. D. Costa, and M. N. Do, "A retina-based perceptually lossless limit and a Gaussian foveation scheme with loss control," *IEEE J. Sel. Topics. Signal Process*, vol. 8, no. 3, pp. 438–453, Jun. 2014.
[29] B. Li, R. Yang, and H. X. Jiang. "Remote-sensing image compression using two-dimensional oriented wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 236–250, Jan. 2011.
[30] A. Karami, M. Yazdi, and G. Mercier. "Compression of hyperspectral images using discrete wavelet transform and Tucker decomposition," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 444–450, Apr. 2012.
[31] X. Zhan, R. Zhang, D. Yin, and A. Z. Hu. "Remote sensing image compression based on double-sparsity dictionary learning and universal trellis coded quantization," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 1665–1669.
[32] C. Jiang, H. Y. Zhang, H. F. Shen, and L. P. Zhang. "Two-step sparse coding for the pan-sharpening of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 7, no. 5, pp. 1792–1805, May 2014.
[33] P. Kulkarni, A. Bilgin, M. W. Marcellin, and J. C. Dagher, "Compression of earth science data with JPEG2000," in *Proc. Hyperspectral Data Compression*, 2006, pp. 347–378.
[34] F. García-Vílchez and J. Serra-Sagristà, "Extending the CCSDS recommendation for image data compression for remote sensing scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3431–3445, Oct. 2009.
[35] K. K. Huang and D. Q. Dai, "A new on-board image codec based on binary tree with adaptive scanning order in scan-based mode," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3737–3750, Oct. 2012.
[36] Z. Wang and A. C. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1397–1410, Oct. 2001.
[37] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," *Proc. SPIE*, vol. 3299, pp. 1–13, 1998.
[38] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor. "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1164–1175, Aug. 1997.
[39] N. Memon, D. L. Neuhoff, and S. Shende, "An analysis of some common scanning techniques for lossless image coding," *IEEE Trans. Image Process.*, vol. 9, no. 11, pp. 1837–1848, Nov. 2000.
[40] C. A. Shaffer, R. Juvvadi, and L. S. Health, "Generalized comparison of quadtree and bintree storage requirements," *Image Vis. Comput.*, vol. 11, no. 7, pp. 402–412, 1993.
[41] D. M. Chandler and S. S. Hemami. "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.
[42] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.
[43] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[44] CCSDS reference test image set, Apr. 2007. [Online]. Available: http://cwe.ccsds.org/sls/docs/sls-dc/
[45] H. Lee and S. Lee, "Visual entropy gain for wavelet image coding," *IEEE Signal Process. Lett.*, vol. 13, no. 9, pp. 553–556, Sep. 2006.
[46] W. J. Zeng, S. Daly, and S. Lei, "An overview of the visual optimization tools in JPEG 2000," *Signal Process.: Image Commun.*, vol. 17, no. 1, pp. 85–104, Jan. 2002.
[47] X. Huang, X. B. Liu, and L. P. Zhang, "A multichannel gray level co-occurrence matrix for multi/hyperspectral image texture representation," *Remote Sens.*, vol. 6, no. 9, pp. 8424–8445, 2014.
[48] S. Saha, "Image compression—From DCT to wavelets: A review," *Crossroads*, vol. 6, no. 3, pp. 12–21, 2000.
[49] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 244–250, Apr. 1992.
[50] V. Bruni and D. Vitulano, "An improvement of kernel-based object tracking based on human perception," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 11, pp. 1474–1485, Nov. 2014.
[51] X. Long, S. N. Li, K. N. Ngan, and L. Ma, "Consistent visual quality control in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 975–989, Jun. 2013.
[52] J. J. Wu, W. S. Lin, G. M. Shi, and A. M. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, Jan. 2013.
[53] K. F. Zhu, C. Q. Li, V. Asari, and D. Saupe, "No-reference video quality assessment based on artifact measurement and statistical analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 533–546, Apr. 2015.



Cuiping Shi (S'13) received the B.S. degree from the Daqing Petroleum Institute, Daqing, China, in 2004 and the M.S. degree from the Yangzhou University, Yangzhou, China, in 2007. She is currently working toward the Ph.D. degree in signal and information processing with Harbin Institute of Technology, Harbin, China.

Her research interests include remote sensing data processing and hyperspectral image processing.

Junping Zhang (M'05) received the B.S. degree in biomedical engineering and instrument from the Harbin Engineering University and Harbin Medical University, Harbin, China, in 1993 and the M.S. and Ph.D. degrees in signal and information processing from Harbin Institute of Technology (HIT), Harbin, in 1998 and 2002, respectively.

She is currently a Professor with the Department of Information Engineering, School of Electronics and Information Engineering, HIT. Her research interests include hyperspectral data analysis and image processing, multisource information fusion, and pattern recognition and classification.

Ye Zhang (M'10) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1982, 1985, and 1996, respectively.

From September 1998 to September 1999, he was a Visiting Scholar at the University of Texas at San Antonio, San Antonio, TX, USA. Since 1985, he has been with the Department of Electrical and Communication Engineering, Harbin Institute of Technology, where he is currently a Professor with the Department of Information Engineering, School of Electronics and Information Engineering. His main research interests include hyperspectral image analysis and processing, image and video compression and transmission, and multisensor-based remote sensing image coprocessing and applications.

A Novel Vision-Based Adaptive Scanning for the Compression of Remote Sensing Images

作者: Shi, CP (Shi, Cuiping)[1,2]; Zhang, JP (Zhang, Junping)[3]; Zhang, Y (Zhang, Ye)[3]

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
卷: 54 期: 3 页: 1336-1348
DOI: 10.1109/TGRS.2015.2478145
出版年: MAR 2016
查看期刊信息

摘要

Most of the compression methods for remote sensing images are often designed under the guidance of mean square error. However, for the vision-related applications, high peaksignal- to-noise ratio (PSNR) does notmean good visual quality. On the other hand, existing compression methods that considering the human visual system (HVS) are usually designed for natural images, without taking the unique characteristics of remote sensing images into account. Focusing on this problem, we present a novel HVS-based adaptive scanning (HAS) scheme for the compression of remote sensing images. First, after the wavelet transform, a retina-based visual sensitivity model is established, and then, the visual weighting mask is generated. Second, for the weighted transformed image, an adaptive scanning method is proposed, which provides different scanning orders among subbands and within a subband, respectively. The former focuses on organizing the codestream according to the importance of weighted subbands, and the latter aims at preserving the direction information of an image as much as possible. Finally, the binary tree codec is utilized. Experimental results show that, as compared with other scan-based compression methods, the proposed HAS-based compression method can provide better visual quality, which makes it more desirable in vision-related applications for remote sensing images.

关键词

作者关键词: Adaptive scanning; embedded coding; human visual system (HVS); remote sensing image compression

KeyWords Plus: PERCEPTUAL-DISTORTION CONTROL; WAVELET TRANSFORM; SPIHT ALGORITHM; QUANTIZATION; JPEG2000; CODEC; EFFICIENT; NOISE; TREE

作者信息

通讯作者地址: Shi, CP (通讯作者)
+ Harbin Inst Technol, Dept Informat Engn, Harbin 150001, Peoples R China.

通讯作者地址: Shi, CP (通讯作者)
+ Qiqihar Univ, Dept Commun Engn, Qiqihar 161000, Peoples R China.

通讯作者地址: Zhang, JP; Zhang, Y (通讯作者)
+ Harbin Inst Technol, Dept Informat Engn, Harbin 150001, Peoples R China.

地址:
+ [1] Harbin Inst Technol, Dept Informat Engn, Harbin 150001, Peoples R China
+ [2] Qiqihar Univ, Dept Commun Engn, Qiqihar 161000, Peoples R China
+ [3] Harbin Inst Technol, Dept Informat Engn, Harbin 150001, Peoples R China

电子邮件地址: scp1980@126.com; zhangjp@hit.edu.cn; zhye@hit.edu.cn

基金资助致谢

基金资助机构	授权号
National Natural Science Foundation of China	61271348 41301455
863 High Technology Program of China	2012AA12A405
	1254CGZH04

引文网络

Prophase Cultivation Project of Technology Achievements Industrialization of Universities in Heilongjiang Province of China

查看基金资助信息

出版商

IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, 445 HOES LANE, PISCATAWAY, NJ 08855-4141 USA

类别 / 分类

研究方向: Geochemistry & Geophysics; Engineering; Remote Sensing; Imaging Science & Photographic Technology

Web of Science 类别: Geochemistry & Geophysics; Engineering, Electrical & Electronic; Remote Sensing; Imaging Science & Photographic Technology

文献信息

文献类型: Article

语种: English

入藏号: WOS:000372369400008

ISSN: 0196-2892

eISSN: 1558-0644

期刊信息

Impact Factor (影响因子): Journal Citation Reports®

其他信息

IDS 号: DG8YL

Web of Science 核心合集中的 "引用的参考文献": 53

Web of Science 核心合集中的 "被引频次": 0

SCI 收录

哈尔滨工业大学 科技文献查新专用章

第 2 条，共 33 条