Scaling-based energy-quality multilevel control for aerial imagery*

GONG Xiao-hui (宫晓蕙)**, LIU Hao** (刘浩)****, SUN Jia-tong** (孙嘉瞳)**, ZHANG Xin-sheng** (张鑫生)**, and SUN Xiao-fan** (孙晓帆)

College of Information Science and Technology, Donghua University, Shanghai 201620, China

(Received 18 August 2017; Revised 1 April 2018)

©Tianjin University of Technology and Springer-Verlag GmbH Germany, part of Springer Nature 2018

This paper designs an energy-quality multilevel framework for the coding and transmission of aerial images, and then introduces a scaling-based intra encoder with flexible sampling factor (*SF*) and quantization parameter (*QP*). By experimentally investigating how different coding configurations affect the complexity-rate-quality characteristics of aerial images, this paper derives a configuration estimation model between energy-quality level and appropriate (*SF*, *QP*) configuration. By utilizing the model, a bivariate control scheme is proposed so as to progressively adjust sender's energy consumption under quality constraints. The experimental results show that the proposed scheme can achieve better energy-quality tradeoff with a wider quality range, and reduce the energy consumption above a certain quality.

Document code: A **Article ID:** 1673-1905(2018)05-0384-7

DOI https://doi.org/10.1007/s11801-018-7190-2

The quality-complexity control of image/video coding is a critical problem in power-constrained imagery applications. Under complexity-rate-quality constraints, the coding mechanism with flexible scaling may yield the enhanced energy-quality performance than that with fixed scaling. The scaling-based coding approaches may be classified into two categories: multi-frame coding and still image coding. The empirical results in the case of multi-frame coding[1-4] show the existence of an optimal scaling policy that achieves the maximal quality for a certain energy. However, these scaling-based coding methods depend on the inter-frame correlation, which cannot be directly applied to still images.

In the case of still image coding, the scaling optimization of JPEG compression has been studied[5]. It has been shown that with a certain coding rate, an optimal sampling ratio exists for the best compression quality, but no automatic ratio determination algorithm has been proposed. If a fixed scaling ratio is known during encoding, the perceptual quality can be improved after super-resolution-assisted decoding[6]. Currently, the joint video team of ISO/IEC and ITU-T has developed the high efficiency video coding (HEVC) standard[7]. HEVC intra encoder has the best compression performance in the current image codecs, while the complexity of HEVC intra encoder is high. Based on the HEVC intra encoder, the better portable graphics (BPG) format possibly replaces JPEG. When the HEVC intra coding mechanism becomes sophisticated, the directrate-distortion modeling becomes a very difficult task. What's more, the above scaling-based coding approaches did not take whether the compression can adapt to transmission power constraints into account. Therefore, it is necessary to jointly consider the rate, complexity and quality for HEVC intra coding.

As a new type of power-constrained imagery applications, aerial images need to be acquired, compressed and transmitted for further analysis. The large amounts of aerial images have to face such problems as high coding and transmission cost. Since the content characteristics of an aerial image cannot be obtained in advance, the content-adaptive source-channel joint coding mechanism is difficultly employed in massive aerial imagery[8], so the uniform quantization and scaling policy for each image have a more general meaning. The computational burden at the sender may hinder the widespread adoption of aerial imagery[9]. An important property of aerial image is that plenty of objects and architectures are manmade. Aerial images generally are classified as a texture-rich image type, and the encoding consumption of aerial image often exceeds that of natural image[10]. At the sender of aerial imagery, the energy consumption and compression quality are two key indexes which are closely related with the coding configuration.

It is a challenging problem how to reduce the energy consumption under quality constraints. One major motivation of this work is from the following observation: for an expected quality, the sender may have spent too many energy resources in encoding and

* This paper was presented in Abstract at the CCF Chinese Conference on Computer Vision, Tianjin, 2017, and recommended by the program committee. This paper has been supported by the Natural Science Foundation of Shanghai (No18ZR1400300), and the Fundamental Research Funds for the Central Universities (No.2232018D3-17).

* E-mail: liuhao@dhu.edu.cn

transmitting the image samples whose quality is much higher than that needed for the receiver. This paper experimentally investigates the impact of different coding configurations on complexity, rate and quality. The tradeoff between energy consumption and image quality can thus be exploited. By deriving a configuration estimation model, a bivariate control scheme is developed to pre-determine the appropriate parameters for any aerial image so as to improve the energy-quality multilevel performance for aerial imagery applications.

At a sender of aerial imagery, image coding and transmission consume a major portion of total energy. Generally speaking, the sender's energy consumption mainly contains two components: the encoding energy for acquiring and processing images, and the transmission energy for transmitting the data to a remote receiver. The encoding energy is a function of computational complexity, and the transmission energy is related to the amount of transmitted data. In this paper, the subscript C denotes an energy-quality multilevel control policy. With the control policy C, T_C denotes its computational complexity in execution time, R_C denotes its transmission rate in kbit/s, and Q_C denotes its compression quality in *PSNR*. For a given Q_C, the sender's energy consumption E_C may be expressed as:

$$E_C=f(T_C)+g(R_C), \tag{1}$$

where $f(\cdot)$ denotes the encoding energy as a function of execution time, and $g(\cdot)$ denotes the transmission energy as a function of transmission rate. Thus with the control policy C, the sender may provide an energy-quality level (E_C, Q_C). For any aerial image, the encoding power is relatively constant over time, which may be accumulated by keeping track of the execution time[11]. Therefore, the encoding energy can be estimated by multiplying the encoding power P_{enc} with its execution time T_C

$$f(T_C)=P_{\text{enc}}\cdot T_C \ , \tag{2}$$

where P_{enc} is a constant in J/ms. During image encoding, a processor almost operates at full load, so the encoding power is close to thermal design power (TDP) of the processor. On the other hand, in terms of data transmission, it is obvious that more energy is required when the transmission rate increases. The transmission energy presents a linear relationship with the transmission rate. Here, E_{tr} denotes the average transmission energy per kbit/s, which depends on transmission distance and path loss index. Therefore, the transmission energy can be estimated by:

$$g(R_C)=E_{\text{tr}}\cdot R_C \ . \tag{3}$$

As given in Eq.(1), the sender's energy consumption can be further formulated as:

$$E_C=P_{\text{enc}}\cdot T_C+E_{\text{tr}}\cdot R_C=P_{\text{enc}}\cdot(T_C+\lambda\cdot R_C), \tag{4}$$

where $\lambda=E_{\text{tr}}/P_{\text{enc}}$ denotes the transmission-coding energy ratio. To obtain multiple energy-quality levels, the existing control schemes only adjust the quantization parameter (*QP*) while maintaining a fixed sampling factor (*SF*). Thus, the number of energy-quality levels is equal to the number of feasible *QP* values. In this paper, the energy-quality multilevel framework provides 52 energy-quality levels. Without loss of generality, the optimization objective of energy-quality multilevel control is to reduce the energy consumption E_C subject to a certain quality Q_C, and simultaneously provide a wider quality range.

Besides the *QP*, the *SF* is another important coding configuration in power-constrained aerial imagery. In this paper, a scaling-based HEVC intra encoder is introduced so as to obtain the appropriate energy-quality multilevel control. At the sender, an original frame is firstly downsampled with a certain *SF*, and then compressed by an HEVC intra encoder. At the receiver, the image is decoded, and then upsampled to the original resolution for further analysis. Fig.1 shows the system diagram of scaling-based aerial imagery, where the original image $\boldsymbol{X}$ is firstly pre-processed by a configuration controller so as to obtain the side information such as *SF* and *QP*. With a given *SF*, the original image $\boldsymbol{X}$ is downsampled to the image $\boldsymbol{X}_{\text{d}}$. Then, the HEVC intra encoder compresses $\boldsymbol{X}_{\text{d}}$ with a given *QP* that meets the rate constraint. The bitstream generator can encapsulate the compressed data and side information into the bitstream $\boldsymbol{Y}$, and then transmits the bitstream to the remote receiver by transmission channel. At the receiver, the bitstream $\boldsymbol{Y}$ is decoded to the image $\boldsymbol{X}'$, and then upsampled to the image $\boldsymbol{X}'_{\text{u}}$ at the original resolution. The sender needs to select the appropriate control policy to reduce the sender's energy consumption while ensuring a certain quality.

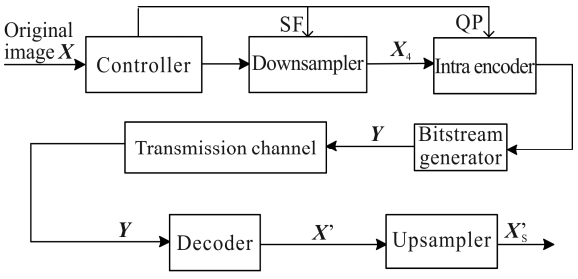


Fig.1 System diagram for scaling-based aerial imagery

The downsampling operation can shrink the original image $\boldsymbol{X}$ from $W\cdot H$ pixels to $W_{\text{d}}\cdot H_{\text{d}}$ pixels. Thus, the downsampled image $\boldsymbol{X}_{\text{d}}$ is obtained at the $W_{\text{d}}\cdot H_{\text{d}}$ resolution. In most cases, the least coding unit is 16×16 pixels. If (W, H) denotes the size of the original image $\boldsymbol{X}$, the size of its downsampled image $\boldsymbol{X}_{\text{d}}$ may be represented as follows:

$$(W_{\text{d}},\ H_{\text{d}})=(W-16\times SF,\ H-16\times SF) \ . \tag{5}$$

After decoding, the upsampling operation can expand the image $\boldsymbol{X}'$ back to $W\cdot H$ pixels. The coding configuration should be determined before employing the

scaling-based HEVC intra encoder. Under such conditions, both encoding energy and transmission energy depend on the *SF* and *QP*. The appropriate coding configuration will be found by heuristically performing the scaling-based HEVC intra encoder with all feasible (*SF*, *QP*) configurations. Based on Eq.(4), the following subsections will experimentally investigate the impact of different coding configurations on the rate, complexity and quality of aerial images, and then derive a bivariate control scheme with a configuration estimation model.

Based on the scaling-based HEVC intra encoder, we will encode all training images to analyze the complexity-rate-quality characteristics of aerial images. We consider that *SF* varies in the range of {1, 2, ..., 49, 50}, and *QP* varies in the range of {0, 1, ..., 50, 51}. Thus, we have 50´ 52 feasible (*SF*, *QP*) configurations. The average performance of all training images is obtained by respectively testing each (*SF*, *QP*) configuration. In terms of complexity characteristics of aerial image, we measure the average execution time in millisecond (ms), which can be translated into the encoding energy by using Eq.(2). Since *SF* and *QP* independently tune configurations, it is more easily to illustrate one configuration while fixing the other. In Fig.2(a), these curves with different markers are generated from the HEVC intra encoder by changing *QP*. In Fig.2(b), for a given *QP*, the relationship between the execution time and *SF* presents the approximately exponential behavior. As can be seen from the two figures, the execution time will decrease when the *QP* or *SF* value increases.

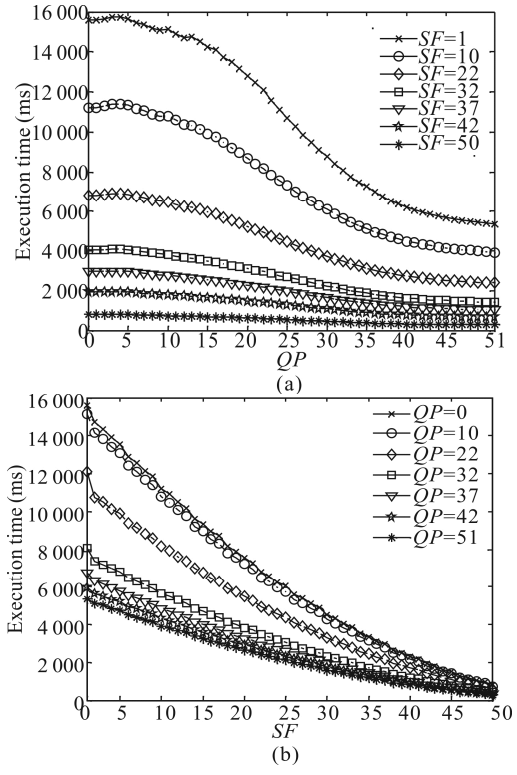


Fig.2 Complexity analysis: (a) Execution time vs. *QP*; (b) Execution time vs. *SF*

In terms of rate characteristics of aerial image, we measure the average transmission rate in kbit/s, which can be translated into the transmission energy by using Eq.(3). The average rates of all training images are obtained by respectively testing each (*SF*, *QP*) configuration. With logarithmic *y*-axis, Fig.3(a) shows the average transmission rate per image vs. *QP* with different *SF*, where the rate and *QP* present the approximately exponential dependence for a given *SF*. With logarithmic *y*-axis, Fig.3(b) shows the average transmission rate per image vs. *SF* with different *QP*, where the rate and *SF* present the approximately linear dependence for a given *QP*. As can be seen from the two figures, the transmission rate will decrease when the *QP* or *SF* value increases.

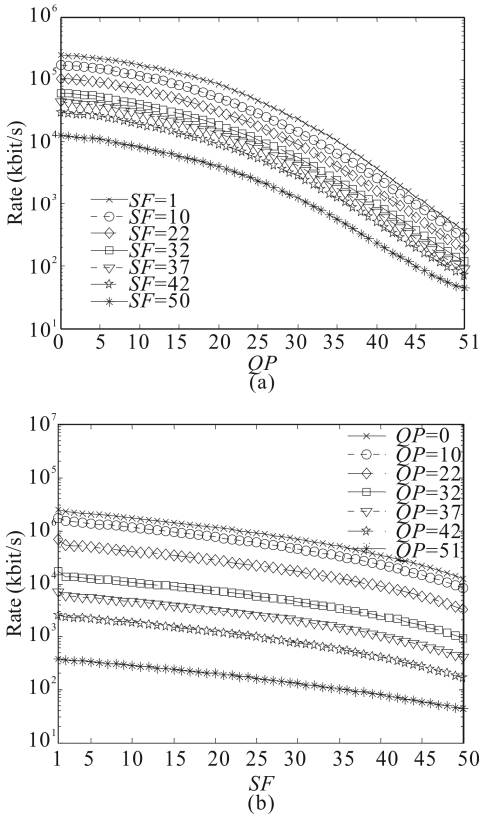


Fig.3 Rate analysis: (a) Transmission rate vs. *QP*; (b) Transmission rate vs. *SF*

In terms of quality characteristics of aerial images, we measure the average compression quality in *PSNR*. When *SF*=1 and *QP*=0, the average *PSNR* of all training images should be equal to its maximum value. Fig.4(a) shows the average *PSNR* per image vs. *QP* with different *SF*, where a family of curves are plotted by changing the *QP* with a given *SF*. If the *SF* value is fixed, the *PSNR* will decrease when the *QP* value increases, although the *PSNR* maintains a certain stability under *QP*=22. Fig.4(b) shows the average *PSNR* per image vs. *SF* with different *QP*, and these curves are generated by changing *SF* with a given *QP*. If the *QP* value is fixed, the *PSNR* will

approximately linearly decrease when the *SF* value increases.

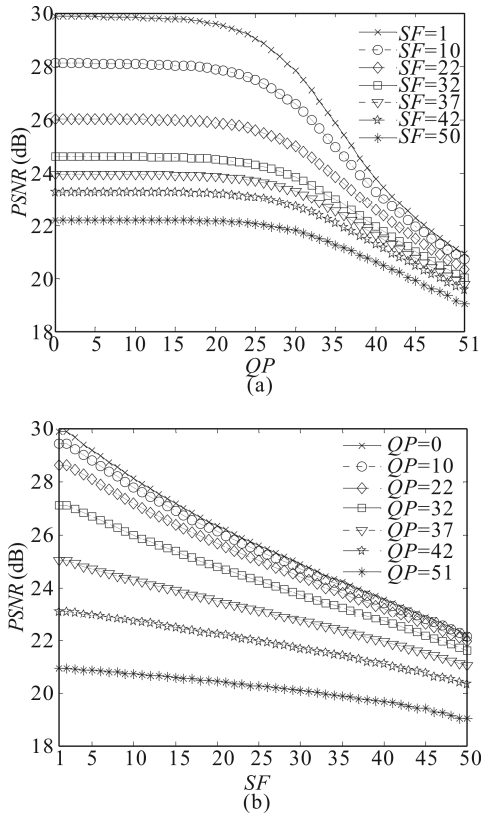


Fig.4 Quality analysis: (a) *PSNR* vs. *QP*; (b) *PSNR* vs. *SF*

To sum up, larger *QP* leads to less execution time, smaller rate and lower quality. Besides, larger *SF* also presents the similar behavior. As can be seen, these figures from Fig.2 to Fig.4 have significantly different variation trends, which can be further used to implement the double-parameter energy-quality optimization. With a higher energy-quality level from 1 to 52, better quality may be obtained at the cost of increased energy consumption. Based on the above experimental results, the next subsection will analyze the statistical relationship between energy-quality levels and coding configurations, and then develop a model-guided bivariate control scheme for determining the appropriate *SF* and *QP* configuration for testing images. The proposed scheme may reduce sender's energy consumption under quality constraints.

The scaling-based HEVC intra encoder executes various (*SF*, *QP*) configurations and generates a complexity-rate-quality space of all training images. In order to obtain a more general conclusion, the transmission-coding energy ratio λ is further extended, which varies from 0.02 to 50 with step size of 0.02. By heuristically feeding the average rate and execution time into the objective function in Eq.(4), the distribution of optimal (*SF*, *QP*) may be obtained by exhaustively searching the minimum energy consumption subject to a certain quality.

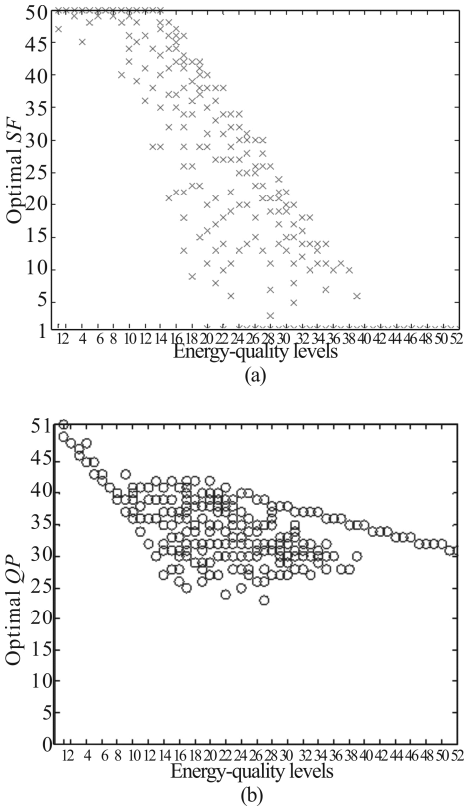


Fig.5 The distributions of optimal configurations at different energy-quality levels: (a) Distribution of optimal *SF* values; (b) Distribution of optimal *QP* values

By analyzing the distribution of optimal *SF* and *QP* values, we will derive a configuration estimation model between energy-quality level and appropriate (*SF*, *QP*) configuration, and the model consists of two relationship functions: the *SF*-level function and the *QP*-level function. Although the complexity-rate-quality characteristics of various images are slightly different, the optimal *SF* or *QP* values distribute within a small range at each energy-quality level. Since the variation is limited in a small range, it is possible to derive an analytic function for the family of mean *SF* or mean *QP* values at different energy-quality levels. Based on the MATLAB cftool toolbox, the mean of optimal *SF* or *QP* values at each energy-quality level is used for the curve fitting process to find the statistical relationship between energy-quality level and (*SF*, *QP*) configuration. The fitting goodness is quantified by the correlation coefficient (*R*-square). In Fig.6, a dot denotes the mean of optimal *SF* or *QP* values at a certain level. In the goodness of fitting, the *R*-square of the *SF*-level function is 0.994 3, and that of the *QP*-level function is 0.988 9. As can be seen from each curve, the piecewise fitting function is acceptable in *R*-square, and the turning points are set to ensure that the function changes monotonically with the level.

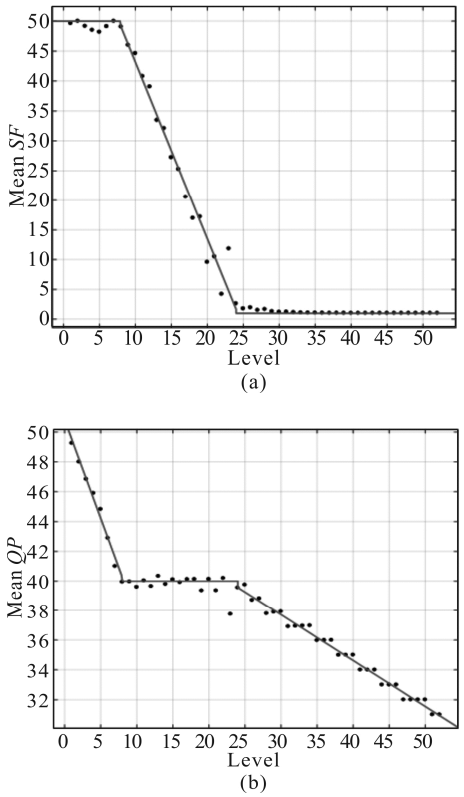


Fig.6 The curve fitting process: (a) *SF*-level function; (b) *QP*-level function

By fitting the distribution of mean *SF* or mean *QP* values, we may derive the following configuration estimation model. Based on linear regression, the function parameters of curve fitting are estimated from the energy-quality statistics of training images. Considering the piecewise linear relationship, the configuration estimation model can be represented as follows

$$SF=\begin{cases}50 & (1\le l<8)\\ \text{round}(72.74-2.962\times l) & (8\le l\le 24)\\ 1 & (24<l\le 52)\end{cases}, \quad (6)$$

$$QP=\begin{cases}\text{round}(50.82-1.322\times l) & (1\le l<8)\\ 40 & (8\le l\le 24)\\ \text{round}(46.96-0.308\,3\times l) & (24<l\le 52)\end{cases}, \quad (7)$$

where l denotes an energy-quality level. After deriving the configuration estimation model, the training images are completely disregarded and the model is then used for the testing images. With the model-guided *SF* and *QP*, the bivariate control scheme can preprocess the testing images before they are fed to the scaling-based HEVC intra encoder. The model-guided bivariate control scheme can progressively adjust sender's energy consumption above a certain quality, and improve the energy-quality scalability.

The proposed bivariate control scheme is evaluated in this section. All experiments are executed on a DELL Tower5810 platform which contains Xeon E5-1603V3 processor, Ubuntu 14.04 LTS (64 bit) operating system and MATLAB R2014a software. The experimental aerial images are all from the USC-SIPI dataset[12], and they are uniformed into 1 024 pixels´1 024 pixels with ITU BT709 color space, where the 40% aerial images (2.2.01~2.2.10) are selected for training and modeling, and the rest (2.2.11~2.2.24) for performance testing and validation. When deriving the configuration estimation model, the training images are respectively encoded by changing the *SF* or *QP* values. The available *SF* values are {1, 2, 3, …,49, 50}, and the available *QP* values are {0, 1, 2, …, 50, 51}. For a given *SF*, the output pixels are obtained by applying the bicubic interpolation on neighboring pixels. When changing the size of an image, a sinc-like anti-aliasing filter is applied to limit the aliasing effect. In fact, the P_{enc} value doesn't affect model-guided (*SF*, *QP*) solution. Here the TDP of the E5-1603V3 processor is 140 W. Without loss of generality, the encoding power P_{enc} is set to 0.14 J/ms. The transmission-coding energy ratio λ varies from 0.02 to 50 with step size of 0.02. The scaling-based HEVC intra encoder is based on the test model reference software HM-16.7[13]. The bivariate control scheme is compared with two existing control schemes: conventional full-size scheme (*SF*=1)[14] and a fixed half-size scheme (*SF*=32). To reduce the energy consumption, the bivariate scheme utilizes the configuration estimation model to determine the appropriate (*SF*, *QP*) configuration.

After the scaling-based HEVC intra encoder processes all testing images with different coding configurations, Fig.7 and Fig.8 give some subjective comparisons of image reconstruction for the image 2.2.12 and the image 2.2.24, where the reconstructed images are based on typical (*SF*, *QP*) configurations. If one coding configuration (*SF* or *QP*) is fixed, the subjective quality will decrease when the other coding configuration (*QP* or *SF*) increases. These subjective quality comparisons show that the scaling-based HEVC intra encoder can utilize different coding configurations to obtain similar quality but far different energy consumption. Therefore, it is possible to further optimize the energy-quality performance.

Due to appropriate image resizing, the energy consumption may benefit from the scaling-based HEVC intra encoder, which is especially suitable for power-constrained aerial imagery. The conventional schemes generate different energy-quality levels by only changing *QP*, while the proposed bivariate scheme utilizes the (*SF*, *QP*) configuration to generate an equal number of energy-quality levels. As the encoder scales down its energy consumption, the compression quality will degrade. With three typical λ cases, Fig.9 shows the energy consumption vs. average quality for three control schemes, where the energy consumption is expressed

with logarithmic coordinate, and a marker point denotes an energy-quality level. When the *PSNR* of reconstructed image increases, its energy consumption will also increases. The conventional full-size scheme performs poorly when the *PSNR* is high, because a large amount of energy is wasted. The fixed half-size scheme performs better at low *PSNR* but the available *PSNR* range is very narrow. Along the *x*-axis, it can be observed that when the *PSNR* increases, the bivariate control scheme can provide a wider quality range and lower energy consumption. The energy saving percentage depends on the energy-quality level. On the whole, these curves show that the bivariate scheme can obtain better energy-quality multilevel control performance than the conventional schemes.

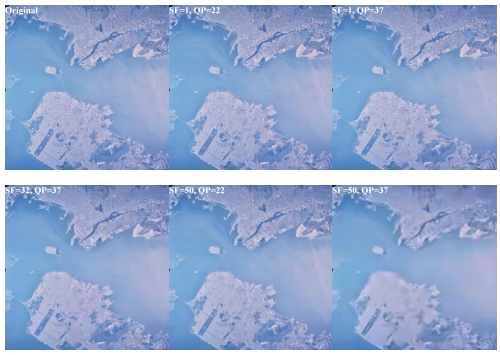


Fig.7 The reconstruction comparison of different (*SF*, *QP*) configurations for the image 2.2.12

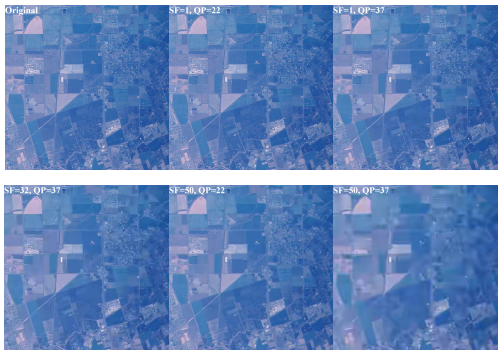
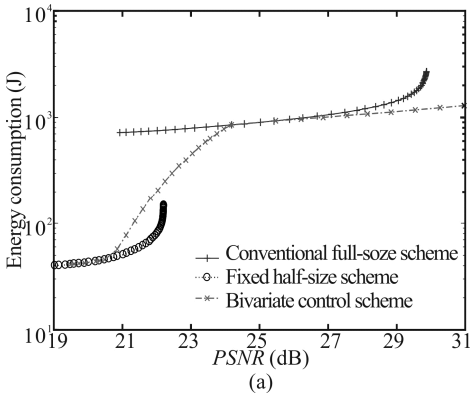


Fig.8 The reconstruction comparison of different (*SF*, *QP*) configurations for the image 2.2.24



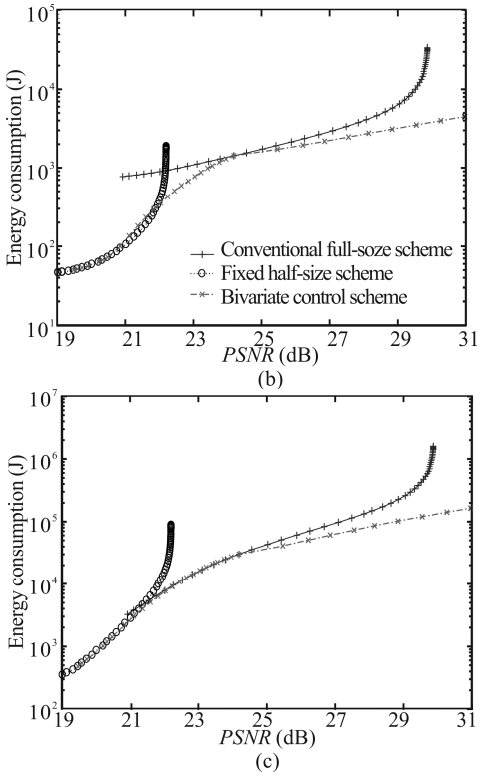


Fig.9 Energy consumption vs. quality for three control schemes with typical *λ* cases: (a) *λ*=0.02; (b) *λ*=1; (c) *λ*=50

This paper designs an energy-quality multilevel framework for the coding and transmission of aerial images, and then introduces a scaling-based HEVC intra encoder with flexible *SF* and *QP* values. By experimentally investigating how different coding configurations affect the complexity-rate-quality characteristics of aerial images, this paper derives the configuration estimation model between energy-quality level and (*SF*, *QP*) configuration. By utilizing the model, this paper proposes the bivariate control scheme so as to progressively adjust sender's energy consumption under quality constraints. The experimental results show that the proposed scheme can significantly reduce the energy consumption above a certain quality while providing a wider quality range, and thus achieve better energy-quality tradeoff than the conventional full-size or half-size schemes.

References

[1] Nguyen V. A., Tan Y. P. and Lin W. S., Adaptive Downsampling/Upsampling for Better Video Compression at Low Bit Rate, Proc. of IEEE International Symposium on Circuits and Systems, 1624 (2008).

[2] Lee H., Lee Y., Lee J., Lee D. and Shin H., IEEE Transactions on Consumer Electronics **55**, 1682 (2009).

[3] Rhee C. E., Kim J. S. and Lee H. J., EURASIP Journal

on Advances in Signal Processing **1**, 1 (2012).

[4] Wang R. J., Huang C. W. and Chang P. C., IEEE Transactions on Circuits and Systems for Video Technology **24**, 1957 (2014).

[5] Bruckstein A., Elad M. and Kimmel R., IEEE Transactions on Image Processing **12**, 1132 (2003).

[6] Barreto D., Alvarez L. D., Molina R., Katsagelos A. K. and Callico G. M., Multidimensional Systems and Signal Processing **18**, 59 (2007).

[7] MPEG-H Part 2, High Efficiency Video Coding (HEVC), ISO/IEC Standard 23008-2, 2013.

[8] Bai H., Lin W., Zhang M., Wang A. and Zhao Y., IEEE Transactions on Circuits and Systems for Video Technology **24**, 1390 (2014).

[9] Pajares G., Photogrammetric Engineering & Remote Sensing **81**, 281 (2015).

[10] Zhang L. M., Han Y. H., Yang Y., Song M. L., Yan S. C. and Tian Q., IEEE Transactions on Image Processing **22**, 5071 (2013).

[11] Redondi A., Baroffio L., Bianchi L., Cesana M. and Tagliasacchi M., IEEE Transactions on Mobile Computing **15**, 3000 (2016).

[12] http://sipi.usc.edu/database/database.php?volume=aerials

[13] HEVC Software Repository—HM-16.7 Reference Model. https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.7.

[14] Pastuszak G. and Abramowski A., IEEE Transactions on Circuits and Systems for Video Technology **26**, 210 (2016).