A Layered Compression Method for High Precision Depth Data

R. Rajkumari[1], Ms. M.Vanitha[2]
[1, 2] Department of ECE
[1, 2] Sriram Engineering college, Perumalpattu, Thiruvallur Tamil Nadu.

Abstract- *In the transmission and depth storage system, we have to determine the compress high precision depth data by using the acquisition technologies. The proposed layered high-precision depth compression framework is used to achieve efficient compression with low complexity in 8-b image/video encoder. In the high precision depth, the depth map is dividing into two layers are the most significant bits (MSBs) and the least significant bits (LSBs). The depth value variation details records by the LSBs layer and the rough depth value provides by the MSBs layer. By proposing the error controllable system in the MSB layer to measure the data correlation of the general depth information sharp edges by taking the quantization error in the MSBs layer to assurance the data format of LSBs. To perform the compression, by using the standard 8-b image/video codec in LSBs layer. Comparing the conventional coding scheme with proposed coding scheme to achieve the good performance the real time depth compression, view synthesis and motion recognition application.*

Keywords- component; Depth compression, layered compression, high precision depth

I. INTRODUCTION

As a typical representation of physical information, depth data has attracted more research interest with respect to the fast development of depth data acquisition technologies and its potential impact on 3D related applications. So far, various depth sensors have been brought forth to record spatial information in the physical world in real-time, such as time of-flight (TOF) cameras [1], stereo cameras [2] and structure light cameras [3]. With the evolution of the depth sensor, the quality of the generated depth map continues to improve the depth precision. For instance, the depth value range of the new depth sensor “Kinect” is up to 4000, and its pixel value is represented by a 12-bit integer in the Windows SDK [3]. TOF camera data also has a depth range of 0 to 2000. As a result, the high precision depth with more than 8 bit depths can easily be accessed by depth sensors for the benefit of depth assisted applications.

The data size of the high precision depth data is larger than the normal depth data. Therefore, an efficient compression scheme is essential for the high precision depth storage and transmission. In some scenarios (e.g. immersive tele presence, remote depth acquisition), depth data has to be transmitted through a network in real-time, which leads to even higher requirements on compression complexity and efficiency.

Depth compression has been studied for years with respect to efficient depth representation and transmission. One straightforward approach is to encode depth map/ sequence using conventional image/video compression schemes [4]. Considering that the depth data contains smooth areas partitioned by sharp edges, with very limited texture, some depth compression approaches [5]–[9] are proposed in an attempt to achieve more efficient depth compression based on the existing image/video coding framework, such as platelet based coding [5], [6] or edge based coding [7]–[9]. In platelet based coding schemes, the depth frame is segmented into regions with different sizes. Each region is encoded by utilizing the homogeneousness of the depth map. For example, Marvon *et al.* [5] segment the depth using a quadtree decomposition, in which each block is modeled using one of three pre-defined piecewise linear functions. Milani and Calvagno [6] partition the depth map based on graph-based image segmentation. In an attempt to reduce the bit cost caused by non-zero high frequency transform coefficients in the edge block, Shen *et al.* [7] propose an edge-adaptive transform, which avoids filtering cross edges by encoding edge positions explicitly. In [8] and [9], shape-adaptive wavelets are employed to ensure that the support for the wavelet lies in the same region separated by edges. Unfortunately, most of these coding schemes are designed for the normal depth data with 8 bit depths and usually implemented based on 8-bit conventional image/video codecs. The data property of high precision depth data is not substantially considered in these schemes.

With the development of depth sensors, appro- aches [10]–[12] for the high precision depth compression are proposed. Mehrotra *et al.* [10] proposed a near lossless depth compression to encode the scaling reciprocals of the Kinect depth values pixel by pixel. Although significant data size reduction is achieved, the strong relationships among the neighboring frames are barely considered. In [12], Fu et al. proposed a Kinect-like depth compression scheme to recover the data correlation destroyed by the noises in temporal and spatial domains.

To fully exploit the data correlation, this scheme is implemented based on the extension version of standard video codec. As we known, with the high profiles defined in fidelity range extension, the standard video codec can offer an extension of bit-depth up to more than 8 bits, such as 14 bits in H.264/AVC and 12 bits in HEVC. However, optimization of encoding implementation at more than 8-bit depths has rarely been achieved. Most optimized implementations of the standard codec are based on the 8-bit depths data format, including software and hardware .Thus, the limitation on the implementation leads to the high computation complexity for the high precision depth compression.

To fully utilize the optimized 8-bit codec, layered compression is an efficient approach for the high precision depth compression. In the traditional high dynamic range (HDR) image/video compression, tone mapping based schemes [13],[14] are widely utilized to convert the high bit-depth image/video to a base layer image/video with 8 bits and an enhancement layer. However, the data format of the high precision depth data is much different from the high dynamic range texture image which is usually represented by floating-point radiance information [15]–[17]. Moreover, the physical meaning of depth map is also different from the texture image. Thus, the HDR coding schemes designed to minimize the visual appearance change during the compression cannot be employed in high precision depth coding. In the multi view image/video coding, layer-based representation is also an efficient approach [18] to exploit the correlation among the multiple views in which the multi view data is partitioned into homogenous layers and each layer is a collection of Epi polar plane image (EPI) lines modeled by a constant depth plane. However, this layered compression scheme is designed for the multiple texture images and cannot be employed in depth data compression. In our previous work [19], we proposed a layered compression scheme leveraging the standard 8-bit codec which can achieve high coding efficiency on Kinect depth. In this paper, we further explore the layered coding framework on the general high precision depth data captured from the distinct depth sensors and optimize the coding performance based on a rate-distortion model.

Based on the characteristics of high precision depth data, we propose a layered compression framework for high precision depth to achieve both high efficiency and low complexity in compression. In this framework, the depth data is first partitioned into two layers in the bit-depth level: the most significant bits (MSBs) layer which contains the multiple highest bit depths with general depth information of the objects and the least significant bits (LSBs) layer which contains the multiple lowest bit depths with tiny variations in depth value on surfaces. In terms of the data properties, distinct coding schemes are designed for each layer. For the MSBs layer, a pixel domain coding scheme is designed to fully exploit spatial correlation among multiple surfaces with sharp edges. For the LSBs layer, a transform-domain coding scheme is employed. An error control algorithm is proposed to guarantee the data format in the LSBs layer is 8-bit, so that the optimized standard image/video codecs can be integrated directly. The experimental results show that the proposed scheme can achieve real-time depth compression with high coding efficiency. Moreover, we investigate the impact of depth compression on the performance of view synthesis and gesture recognition and observe that better performances can be acquired using the depth map reconstructed from the proposed scheme in comparison with the conventional scheme. The main contributions of this framework can be summarized as follows: 1) We propose a layered representation form of high precision depth by simple data partition in the bit-depth level to support a multi-grained description of depth. 2) The layered coding scheme can leverage 8-bit standard codec. Most of 8-bit image/video encoders including software and hardware can be integrated easily to accelerate the coding processing. 3) The proposed coding scheme can achieve high coding efficiency and low coding complexity with desired depth feature preserved for relevant applications and can be integrated into a real-time remote depth acquisition system.

The rest of the paper is organized as follows. In Section II, we provide data analysis on high precision depth data and then propose the layered compression framework. Section III describes the compression scheme of each layer in detail. The optimization of the bit-depth level partition is studied in Section IV. The experimental results are presented in Section V. Section VI concludes this paper.

II. MOTIVATION AND THE PROPOSED FRAMEWORK

In this section, we first provide a data analysis of high precision depth, which indicates the motivation of the pro- posed framework, and then introduce the layered depth coding framework.

A. Motivation Based on Data Analysis

For the high precision depth data, even though the pixel value range is huge for the overall depth map, the local range in most of regions is small. To investigate the data characteristic, we measure the pixel range for each 16×16 block in a depth map by calculating the difference between the maximum and minimum valid depth value. The last two depth sequences are captured at 30fps with a resolution of 800×480 by stereo depth camera. The depth value ranges of these six sequences are all [0, 4096-1] with 12 bit depths. From the results we can see that considerable blocks have a low pixel range. This result implies that depth value usually varies within a small range in the local region. The blocks with large range are likely located along the occlusion boundaries. Even though the depth data has a large pixel value range around the boundary, it is usually composed of several smooth regions located in the foreground and background and the pixel value range is small in each region. This data property implies that the depth value within the block is sparsely distributed and the pixels can be easily classified as several clusters. To investigate this, we use a constrained K-means to classify the pixels within one block and guarantee that the difference between the maximum and minimal depth values in each cluster is within a constraint. The clustering starts from one cluster. If the range of the cluster exceeds the constraint, we increase the cluster number until the constraint is met.

The constraint is set as the form of $2^{l} - 1$ and fours constraint values, which are 255, 127, 63 and 31, are set to perform the investigation. Given the distinct constraint value, we calculate the block proportion with the distinct cluster numbers based on 720000 blocks from six depth sequences .It is obvious that if the constraint is tightened with a smaller difference value, there will be more clusters within one block. Note that using four clusters can represent more than 70% blocks and the quantization error can be bounded within 31. If the high precision depth data has *n* bit depths and highest *m* bit depths are extracted as one layer depth data, can also reflect the clustering results of this depth layer with the maximum quantization error as $2^{l-(n-m)} - 1$ correspondingly precision depth data owns clustering characteristic that will be more obvious for the depth data extracted from the high bit-depth levels.

Edge-adaptive transform, which avoids filtering cross edges by encoding edge positions explicitly. Shape-adaptive wavelets are employed to ensure that the support for the wavelet lies in the same region separated by edges. Unfortunately, most of these coding schemes are designed for the normal depth data with 8 bit depths and usually implemented based on 8-bit conventional image/video codecs. The data property of high precision depth data is not substantially considered in these schemes.

B. Proposed Framework

With respect to the property of the high precision depth data, we propose a layered depth compression framework to partition the depth into two layers in bit-depth level, the MSBs layer and the LSBs layer, with low pixel value ranges to represent the high precision depth. The MSBs layer contains the general depth information and gives a coarse grained description of the objects' surfaces, while the LSBs layer records the remaining detailed depth information to describe the tiny variations on the surfaces. The distinct coding schemes are designed for each layer. The specific implementation of the proposed scheme is depicted.

Since the depth sensor is limited on accuracy and capability, depth value capture is usually failed due to several limitations [20], [21] and the pixels in these regions are invalid and represented as zero in depth map. To remove the sharp boundary caused by the invalid pixels and facilitate block-based coding, the depth frame is first fed into the padding module, where the invalid depth pixels in each block are padded by the average of valid pixel values within the block. Meanwhile, a binary mask map recording invalid depth positions is generated as the side information, which will be entropy encoded and transmitted to the decoder for reconstruction.

After padding, the high precision depth represented by *n*-bit integer is partitioned into two layers in bit-depth level: the MSBs layer and the LSBs layer, initialized by the highest *m*-bit and the lowest remaining *(n-m)*-bit integers respectively shown in Fig. 4. In the MSBs layer, the depth value distributes sparsely with clustering property. Considering the efficiency of the transform-domain coding scheme is inadequate for the image compression with sharp edges, a pixel domain coding scheme is proposed that the MSBs layer depth in one block is represented as several major depth values and the corresponding index map through quantization and compressed by entropy coding. Since MSBs layer contains more significant depth information, the coding distortion is recorded and added back to the LSBs layer highest remaining bit-depth for further encoding.

In the LSBs layer, the depth map is smooth after extracting sharp edge information in MSBs layer and the pixel values change in a continuous manner. Thus, the transform-domain coding scheme is employed for efficient compression. An adaptive quantization scheme with error control is proposed in MSBs layer coding to guarantee the LSBs layer depth data fed to encoder is within 8 bits, so that the existing optimized image/video codecs can be integrated directly. An example of the high precision depth data with layer partition is shown in Fig. 5, in which the highest 8 bits depth data are partitioned into the MSBs layer while the lowest 4 bits are sent to the LSBs layer.

In the decoder, the MSBs/LSBs layer bit streams are decoded respectively. Then the decoded two-layer depth data are merged as a high precision depth map. The reconstructed depth map is generated by merging the depth map with mask map.

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

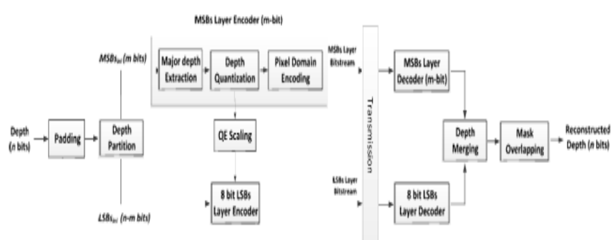

Figure 1. The architecture of the proposed high precision depth encoding and decoding.

III. LAYERED DEPTH COMPRESSION

A. MSBs Layer Compression

The MSBs layer block can be expressed concisely by several major depth values and the corresponding index map through quantization that the pixels with close depth values are clustered as one cluster and one representative depth value is selected to approximate all the pixels in this cluster. Assume that an MSBs block with N pixels can be classified into k clusters. In each cluster S_k, all the pixels are quantized as one major depth value MD_k

1) Histograam Based Quantizatin

In the proposed scheme, the major depth value in each cluster and the corresponding clustering range are determined based on the histogram of MSBs layer block. In order to reduce the quantization distortion. Given the selected major depth, the neighboring depth values in the quantization window are quantized as the major depth. Then the pixel numbers of these depth values including major depth are set as zero in the histogram to indicate that these depth values have been quantized. Then the major depth is selected from the updated histogram and this process is performed several times until K_{max} major depths are selected or all the non-zero depth values in the histogram are clustered. In the proposed scheme, the maximum major depth number K_{max} is set

2) Pixel Domain Encoding

After quantization, the MSBs block data can be represented as the major depths and the corresponding index map. The index values will range from $0\sim K$, in which K is equal to the maximum major Given one major depth, the neighboring depth values in the quantization window are quantized as the major depth. To ensure that the pixel value range of $LSB\,s_q$ is within 8 bits, the window size should be bounded. Assuming that depth data has n bits and the MSBs layer has m bits, there will be n-m bits initially fed to LSBs layer. Then the remaining highest 8-(n-m) bits it LSBs layer can be filled by the quantization error from MSBs layer

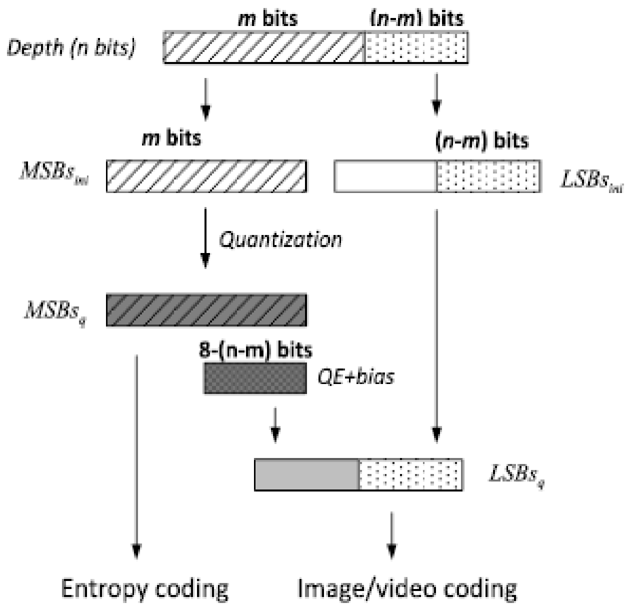

Figure 2. The evolution of bit-depth allocation between two layers

The depth map is smooth after extracting sharp edge information in MSBs layer and the pixel values change in a continuous manner. Thus, the transform-domain coding scheme is employed for efficient compression. An adaptive quantization scheme with error control is proposed in MSBs layer coding to guarantee the LSBs layer depth data fed to encoder is within 8 bits, so that the existing optimized image/video codecs can be integrated directly. An example of the high precision depth data with layer partition in which the highest 8 bits depth data are partitioned into the MSBs layer while the lowest 4 bits are sent to the LSBs layer. In the decoder, the MSBs/LSBs layer bit streams are decoded respectively. Then the decoded two-layer depth data are merged as a high precision depth map. The reconstructed depth map is generated by merging the depth map with mask map

B. LSBs Layer Encoding

In the proposed scheme, after extracting the most significant depth information from the original depth, the sharp edge feature is preserved in MSBs layer, while the generated LSBs layer has depth variation with a smaller range than the original depth. Therefore, a transform-domain image/video coding scheme is employed to further exploit the correlation in spatial and temporal domains.

Due to the error control in the MSBs layer coding, the pixel value range of the LSB sq is within 0∼255. Most state-of-the- art 8-bit codecs can be integrated into the proposed coding structure, such as JPEG, MPEG-2, H.264 and HEVC. The integrated codecs can be chosen in terms of the specified system requirements for coding efficiency and complexity word

IV. BIT-DEPTH LEVEL PARTITION

The bit-depth level partition between two layers determines the distribution of depth information and the coding performance in each layer. In order to optimize the coding efficiency, a frame-level rate distortion model is derived to assist with the bit depth partition between the two layers. Given the quantized MSBs layer encoded in pixel domain without distortion and the lossy compression in LSBs layer, the bit-depth partition problem is a rate-distortion optimization problem to an increase in both bitrate and distortion. In the proposed scheme, the high gradient information is preserved in the MSBs layer and compressed in pixel domain while the remaining smooth region data is fed to the LSBs layer for compression. Therefore, how much high gradient information is preserved in the MSBs layer directly influences the coding performance of the two layers. To evaluate the high gradient information amount, we count the high gradient pixel number within one block. Since two nonadjacent pixels in a histogram of the MSBs will be considered as two distinct objects, the current pixel is treated as one high gradient pixel if the difference between the current pixel and the left top neighboring pixel is beyond an adaptive threshold 2n−m In terms of this rule, the high gradient pixel number within one block HGm is calculated. We randomly select 4800 depth blocks from the six depth sequences in Fig. 1 and investigate the relationship between the high gradient pixel number and MSBs bit number,. When more bits are allocated to the MSBs layer, the depth map in the MSBs layer will be more precise and the high gradient pixel number will increase proportionally.

We capture nine sequences including all four gestures. Gesture recognition is performed based on the original depth map, the reconstructed depth map from HEVC and the reconstructed depth map from the proposed scheme, respectively using the same recognition implementation

The CRR based on the compressed depth data by the proposed scheme is much higher than that of HEVC and close to the result based on the original depth data. This is because the error control algorithm is adopted in the proposed scheme. The MSBs layer compression is lossless while the distortion may only be introduced in the LSBs layer within a limited range. Fig. 18 shows one intermediate processing result in gesture recognition. We can see that there exist artifacts around the boundary of the people and in the background of the screen in the depth map from HEVC reconstruction, which will lead to failures in hand tracking. Notice that the reconstructed depth map by the proposed scheme provides a clear boundary between people and screen background that will benefit gesture recognition.

V. CONCLSION

In this paper, we presented a layered compression scheme for high precision depth data. There are three main advantages to the proposed scheme. First, we utilize a simple partition method at bit-depth level to enable an extended encoding range of a state-of-the-art 8-bit image/video codec for depth data with higher bit depths. Second, an error controllable coding scheme has been designed to better preserve depth features for relevant applications. Third, the proposed coding scheme can achieve real-time encoding and decoding rates with a low bitrate cost and can easily be integrated into a high precision depth data transmission system.

REFERENCES

[1] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor— System description, issues and solutions," in Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW), 2004, p. 35.

[2] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 9, pp. 920–932, Sep. 1994.

[3] Kinect. [Online]. Available: https://dev.windows.com/en-us/kinect, accessed Sep. 26, 2015.

[4] S. Grewatsch and E. Muller, "Evaluation of motion compensation and coding strategies for compression of depth map sequences," in Proc. SPIE 49th Annu. Meeting Opt. Sci. Technol., 2004, pp. 117–124.

[5] Y. Morvan, D. Farin, and P. H. N. de With, "Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images," in Proc. IEEE Int. Conf. Image Process. (ICIP), vol. 5. Sep./Oct. 2007, pp. V-105–V-108.

[6] S. Milani and G. Calvagno, "A depth image coder based on progressive silhouettes," IEEE Signal Process. Lett., vol. 17, no. 8, pp. 711–714, Aug. 2010.

[7] G. Shen, W.-S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," in Proc. Picture Coding Symp. (PCS), 2010, pp. 566–569.

[8] A. Sánchez, G. Shen, and A. Ortega, "Edge-preserving depth-map coding using graph-based wavelets," in Proc. Conf. Rec. 43rd Asilomar Conf. Signals, Syst. Comput., 2009, pp. 578–582.

[9] M. Maitre and M. N. Do, "Depth and depth–color coding using shape- adaptive wavelets," J. Vis. Commun. Image Represent., vol. 21, nos. 5–6, pp. 513–522, 2010.

[10] S. Mehrotra et al., "Low-complexity, near-lossless coding of depth maps from Kinect-like depth cameras," in Proc. IEEE 13th Int. Workshop Multimedia Signal Process. (MMSP), Oct. 2011, pp. 1–6.

[11] C. Wang, S. C. Chan, C. H. Ho, A. L. Liu, and H. Y. Shum, "A real- time image-based rendering and compression system with Kinect depth camera," in Proc. 19th Int. Conf. Digit. Signal Process. (DSP), 2014, pp. 626–630.

[12] J. Fu, D. Miao, W. Yu, S. Wang, Y. Lu, and S. Li, "Kinect-like depth data compression," IEEE Trans. Multimedia, vol. 15, no. 6, pp. 1340–1352, Oct. 2013.

[13] G. Ward and M. Simmons, "Subband encoding of high dynamic range imagery," in Proc. 1st Symp. Appl. Perception Graph. Vis., 2004, pp. 83–90.

[14] Z. Mai, H. Mansour, P. Nasiopoulos, and R. K. Ward, "Visually favorable tone-mapping with high compression performance in bit-depth scalable video coding," IEEE Trans. Multimedia, vol. 15, no. 7, pp. 1503–1518, Nov. 2013.

[15] G. Ward, "Real pixels," in Graphics Gems II. San Diego, CA, USA: Academic, 1991, pp. 80–83.

[16] F. Kainz and R. Bogart. Technical Introduction to OpenEXR. [Online] Available: http://www.openexr.com/TechnicalIntroduction.pdf

[17] G. W. Larson, "LogLuv encoding for full-gamut, high-dynamic range images," J. Graph. Tools, vol. 3, no. 1, pp. 15–31, 1998.

[18] A. Gelman, P. L. Dragotti, and V. Velisavljevic´, "Multiview image coding using depth layers and an

optimized bit allocation," IEEE Trans. Image Process., vol. 21, no. 9, pp. 4092–4105, Sep. 2012.

[19] 19] D. Miao, J. Fu, Y. Lu, S. Li, and C. W. Chen, "Layered compression for high dynamic range depth," in Proc. IEEE Vis. Commun. Image Process. (VCIP), Nov. 2012, pp. 1–6.

[20] K. Khoshelham, "Accuracy analysis of Kinect depth data," in Proc. ISPRS Workshop Laser Scanning, vol. 38. 2011, pp. 133–138.

[21] S.-W. Jung, "A bilateral hole filling algorithm for time-of-flight depth camera," Proc. IS&T/SPIE Electron. Imag. Int. Soc. Opt. Photon., Feb. 2013.

[22] R. M. Gray, "Vector quantization," IEEE ASSP Mag., vol. 1, no. 2, pp. 4–29, Apr. 1984.

[23] X. Wu, "Color quantization by dynamic programming and principal analysis," ACM Trans. Graph., vol. 11, no. 4, pp. 348–372, 1992.

[24] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K -means clustering algorithm," J. Roy. Statist. Soc. C (Appl. Statist.), vol. 28, no. 1, pp. 100–108, 1979.

[25] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," IEEE Trans. Circuits Syst. Video Technol., vol. 7, no. 1, pp. 246–250, Feb. 1997.

[26] G. K. Wallace, "The JPEG still picture compression standard,"Commun.ACM, vol. 34, no. 4, pp. 30–44, 1991.