

# Lossless Compression of JPEG Coded Photo Collections

Hao Wu, Xiaoyan Sun, *Senior Member, IEEE*, Jingyu Yang, Wenjun Zeng, *Fellow, IEEE*  
and Feng Wu, *Fellow, IEEE*

**Abstract**—The explosion of digital photos has posed a significant challenge to photo storage and transmission for both personal devices and cloud platforms. In this paper, we propose a novel lossless compression method to further reduce the size of a set of JPEG coded correlated images without any loss of information. The proposed method jointly removes inter/intra image redundancy in the feature, spatial, and frequency domains. For each collection, we first organize the images into a pseudo video by minimizing the global prediction cost in the feature domain. We then present a hybrid disparity compensation method to better exploit both the global and local correlations among the images in the spatial domain. Furthermore, the redundancy between each compensated signal and the corresponding target image is adaptively reduced in the frequency domain. Experimental results demonstrate the effectiveness of the proposed lossless compression method. Compared to the JPEG coded image collections, our method achieves average bit savings of more than 31%.

**Index Terms**—Image compression, lossless, JPEG, recompression, image set, image collection, image coding

## I. INTRODUCTION

THE increasing number of digital photos on both personal devices and the Internet has posed a big challenge for storage. With the fast development and prevailing use of handheld cameras, cost as well as required photography skills are much lower than before. Users today have gotten used to taking and posting photos profligately with mobile phones, digital cameras, and other portable devices to record daily life, share experiences, and promote businesses. According to recent reports, Instagram users have been posting an average of 55 million photos every day [1]. Facebook users are uploading 350 million photos each day [2]. How to store, backup, and maintain these enormous amount of photos in an efficient way has become an urgent problem.

The most popular way to reduce storage sizes of photos is via JPEG compression [3]. It is designed for reducing the size of photos taken in realistic scenes with smooth variations of tone and color. Though several superior formats such as JPEG

2000 [4] and JPEG XR [5] have been developed subsequently, the JPEG baseline is exclusively used as a common and default format in almost all imaging devices like digital cameras and smart phones. Consequently, the overwhelming majority of images stored in both personal devices and the Internet are in JPEG format. In this paper, we focus on the lossy coded baseline JPEG image, which is referred to as the JPEG coded image in the rest of this paper.

The JPEG baseline, which was first standardized in 1992, nevertheless leaves plenty of room for further compression. It reduces inter-block redundancy among DC coefficients via a differential coding and exploits the statistical correlation inside each block through a table based Huffman coding. The performance of image compression can be enhanced by introducing both advanced intra prediction methods, such as pixel-wise [6] [7] and block-wise [8] [9] intra predictors, and high efficiency entropy coding methods like arithmetic coders [10].

We notice that all the aforementioned intra prediction methods operate on pixels in the raw images (i.e. original captured images without any compression). To further compress JPEG coded images, these methods have to first decode the JPEG coded images to the spatial domain (e.g. YUV or RGB space) and then perform intra or inter predictions. When lossless compression is mandatory as for photo archiving, backup, and sync, the prediction residues may need more bits for lossless compression compared with the corresponding original JPEG coded images (as demonstrated in Table II of Section VIII-A). To address this problem, some lossless compression methods are proposed to reduce the spatial redundancy in the frequency domain [11] or by designing advanced arithmetic entropy coding methods [12]–[16]. Besides, commercial or open-source archivers such as WinZip [17], PackJPG [18], StuffIt [19] and PAQ [16] have employed dedicated algorithms to reduce the file size of individual JPEG coded images.

On the other hand, when dealing with a group of correlated images, the inter image redundancy can be exploited by organizing images as a pseudo sequence and compress the sequence like a video [20]–[28], or subtracting a representative signal (e.g., an average image) from each image [29]–[32] and coding the residues using image coding methods. Recently, inter image redundancy has also been investigated for image compression using a predefined 3D model [33], similar images retrieved from clouds [34]–[36] or videos [37]. However, all these compression schemes are designed for coding pixels in raw images. To the best of our knowledge, there is no lossless compression scheme for the existing JPEG coded image set

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work was supported by the Distinguished Young Scholars Program under Grant 61425026 and the Normal Project under Grant 61372084 from the National Natural Science Foundation of China.

H. Wu and J. Yang are with Tianjin University, Tianjin 300072, China (e-mail: wuzi@tju.edu.cn; yjy@tju.edu.cn)

X. Sun and W. Zeng is with Microsoft Research, Beijing 100080, China (e-mail: xysun@microsoft.com; weizeng@microsoft.com).

F. Wu is with University of Science and Technology of China, Hefei 230027, China (e-mail: fengwu@ustc.edu.cn).

presented before.

In this paper, we propose a novel compression scheme to further compress a set of JPEG coded correlated images without loss. Given a JPEG coded image set, we propose to remove both inter and intra redundancies by a hybrid prediction in the feature, spatial, and frequency domains. We first evaluate the pair-wise correlation between images by introducing the feature-based measurement so as to determine the prediction structure which is robust to scale, rotation, and illumination. The disparity between images is then compensated by both the global (geometric and photometric) alignments and the HEVC-like local motion estimation [9] in the spatial domain. Furthermore, we reduce both the inter and intra redundancies via frequency domain prediction and context-adaptive entropy coding. Compared with our preliminary work reported in [38], we not only provide more details and discussions of our scheme here, but more importantly we further improve the coding performance by introducing both the intra frame lossless compression algorithm and advanced entropy coding methods. Experimental results demonstrate the advantage of our scheme in terms of achieving much higher coding efficiency and lossless representation of JPEG coded files. Our scheme is able to greatly reduce the cost of storage and transmission of JPEG-coded image collections (*e.g.* geo-tagged images and personal albums) transparently for personal and cloud applications.

The rest of this paper is organized as follows. Section II reviews related work. Section III describes the framework of the proposed hybrid coding scheme. The three modules, feature domain determination of prediction structure, spatial domain disparity compensation, and frequency domain redundancy reduction, are presented in Sections IV, V, and VI, respectively. Section VII shows our experimental results. Section VIII concludes this paper.

## II. RELATED WORK

In this section, we first briefly introduce the JPEG image compression scheme and then review related work on both lossless compression of individual JPEG images and image set compression.

### A. Baseline JPEG Compression

The JPEG group has specified a family of image coding standards. The most popular one is the baseline JPEG. Fig. 1 shows the key components of the baseline encoder.

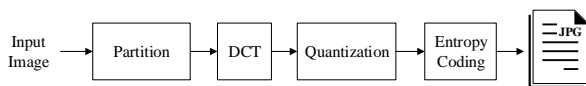


Fig. 1 Illustration of the baseline JPEG encoder.

As shown in this figure, an input image is divided into  $8 \times 8$  blocks. Each block is then converted into a frequency domain by an  $8 \times 8$  DCT, followed by the scalar quantization which

is usually implemented with a set of quantization matrices indexed by a quality factor  $Q \in \{1, 2, \dots, 100\}$ . The quantized DC coefficients are predicted by DPCM (Differential pulse code modulation) while the AC ones are zig-zag scanned before going through the Huffman-based entropy coding.

### B. Lossless Compression of Individual JPEG Images

The most straightforward way to reduce the storage size of a JPEG coded image losslessly is to replace the Huffman coding by an arithmetic coder. In fact, the JPEG extension [39] has already supported an adaptive binary arithmetic coder which can reduce the file size by 8-10%, as reported in [14].

Enhanced performance is achievable by further exploiting inter-block correlation in the intra prediction and designing dedicated entropy coding as well. Ponomarenko *et al.* [12] proposed to separate the quantized DCT coefficients into bit-planes and design context models with regard to the correlations of coefficients within a block, between neighboring blocks and among different color layers. The quantized DCT coefficients can also be re-ordered or grouped based on similar statistical properties, either via a sorting transform in which the order of DCT coefficients are predicted from those of the previous coded blocks [13] or using three scan patterns for low, middle, and high bands [14], respectively, for the adaptive arithmetic coding. Lakhani [15] proposed a new prediction method which derives the one dimensional DCT coefficients of boundary columns/rows from those of adjacent blocks followed by an advanced predictive entropy coding method. Matsuda *et al.* [11] proposed to employ the H.264-like intra prediction method to exploit the inter-block correlations of the quantized DCT coefficients before the adaptive arithmetic coder.

We note that, though effective, all these methods exploit only the redundancy within each individual image. For photo collections that contain correlated images, inter-image redundancy is not exploited in these methods but will be efficiently reduced in our proposed lossless compression scheme.

### C. Image Set Compression

For most digital cameras and smart phones in use today, photography is quite simple and cheap. Taking multiple shots becomes one of the best and general ways to ensure the quality of captured photos, resulting in large numbers of highly correlated image collections in personal photo albums. Such collections may also come from various image sets, *e.g.* geo-tagged photo collections and burst images. When dealing with a group of correlated images, several image set compression schemes have been proposed in the literature. They can be roughly divided into two classes.

The first class of approaches generates a representative signal (RS) (*e.g.*, an average image) from all correlated images to extract the common redundancy among them. Then both the RS and the difference signal between each correlated image and the RS are compressed. Approaches in this class put effort into the RS generation, for example, by the Karhunen-Loeve transform (KLT) [29], the centroid method [30], the max-min differential method [31], the max-min predictive method

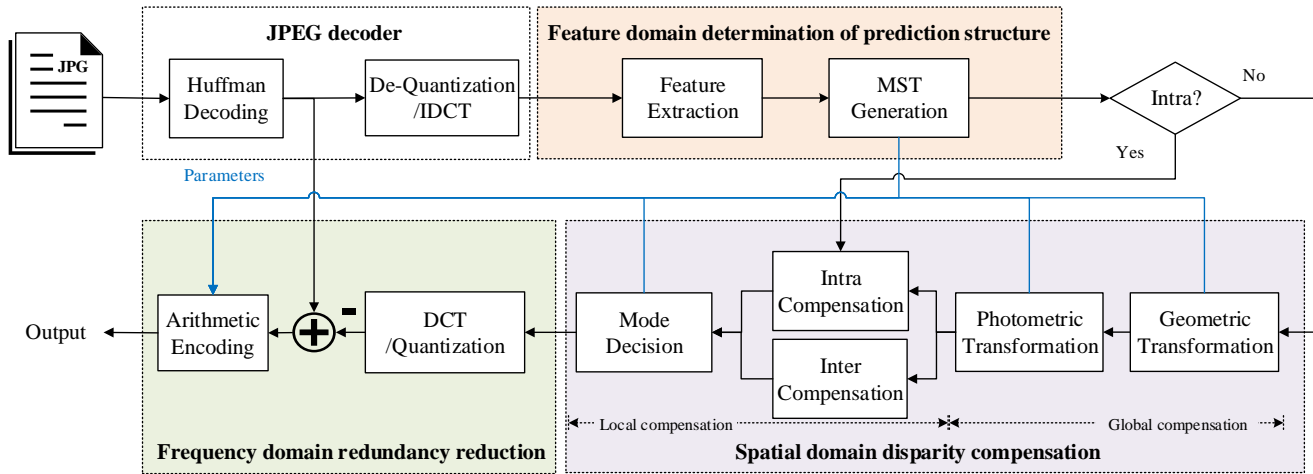


Fig. 2 Architecture of our proposed encoder. For each input JPEG coded image collection, images are first decoded. Then the prediction structure of the collection is determined using local features via MST. For the root image, HEVC-like intra prediction is used to exploit intra correlation. For the other images, the joint global and local disparity compensation is employed to better exploit the correlation between images. Lastly, prediction residues are generated in the frequency domain followed by the context adaptive arithmetic encoder.

[31], and the low-frequency template [32]. Accordingly, these approaches work efficiently when images in a set are similar enough, but share the same limitations when dealing with general image sets. First, they are not robust enough with respect to rotation, scale, and illumination changes which are common in general image sets. Second, they are not sufficiently capable of exploiting the pair-wise correlation between images as different pairs of images may have different correlations.

Rather than set level correlation, the second class of approaches focuses on pair-wise correlations between images. One group of methods finds the optimal coding order of an image set. Schmieder *et al.* evaluate the hierarchical clustering methods for image set compression [20]. Chen *et al.* optimize the coding order of correlated images by minimizing prediction costs [21]. The image set can be clustered by a minimum spanning forest (MSF) and each cluster can be described by a minimum spanning tree (MST) and coded with inter-image prediction. Lu *et al.* apply MPEG-like block-based motion compensation (BMC) to exploit the correlation between images [22]. Au *et al.* introduce a pixel-wise global motion compensation before the BMC [23]. Zou *et al.* apply the advanced BMC in HEVC to image set compression [24]. Though more efficient than RS-based methods, these approaches may lose efficiency when dealing with image sets that have large variations in rotation, scale, and illumination. To address this problem, Shi *et al.* [25] [26] proposed to model the disparity among images with local features and introduced feature-based matching to better exploit the correlations among images in a collection.

On the other hand, all these methods were proposed to compress the pixels in raw images in a lossy way. When extended for lossless compression of JPEG coded images, these methods may not perform well and may even be worse than using the

original JPEG files since they take no consideration of the JPEG coding effects as well as the characteristics.

### III. OVERVIEW OF OUR PROPOSED SCHEME

In this paper, we propose a new coding method for lossless compression of JPEG coded image collections. Specifically, we compress a JPEG coded image collection by making use of both the inter correlation among images and the intra correlation within each image in the feature, spatial, and frequency domains jointly.

Fig. 2 illustrates the architecture of our lossless encoder. For each input JPEG coded image collection (the JPEG icon in Fig.2), we decode all JPEG files before further compression, resulting in the corresponding YUV image set. Then the prediction structure of the image set is determined based on the similarity between each pair of images in the feature domain. The prediction structure is formed in a tree structure generated from a directed graph via the minimum spanning tree (MST) algorithm in which parent nodes (*i.e.* images) can be used as references to predict their children. Section IV gives a detailed description of our feature domain determination of the prediction structure.

Based on the prediction structure, we then exploit both the inter and intra redundancies in the spatial domain. For inter coded images, the disparity between each pair of target and reference images is reduced by joint global and local compensations in the pixel space. Specifically, larger geometric deformations and illumination differences are compensated by the global homography and photometric transforms (illustrated by global compensation in Fig. 2), respectively, while smaller disparities are further compensated by the HEVC-like block-based intra/inter prediction (local compensation in Fig. 2). For the root image in each MST, the global compensation is bypassed and only intra prediction is performed. All the

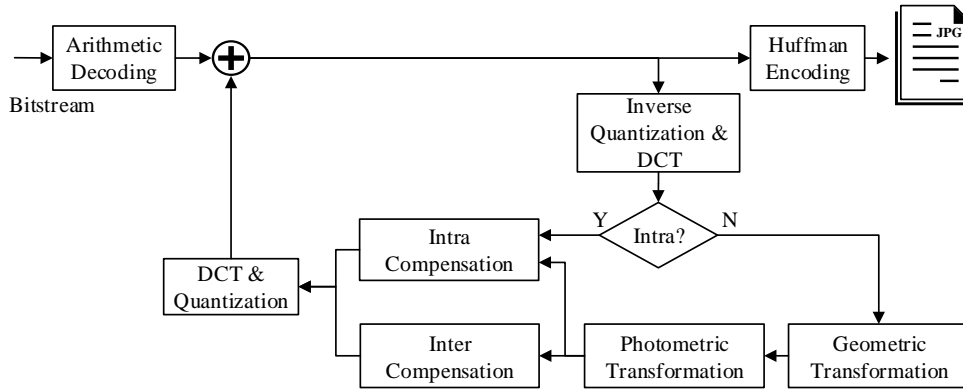


Fig. 3 Architecture of our proposed decoder. The input bitstream is first entropy decoded, resulting in the residues. For intra-coded images/blocks, the intra compensation is performed followed by DCT transformation and quantization. The quantized signal is then added to the residues, generating quantized DCT coefficients of original JPEG images. At last, these coefficients go through JPEG entropy coding to recover JPEG binary files. For inter coded images/blocks, both the global and local compensations are further involved in the decoding process.

parameters of MST, transformations, and modes are entropy coded and stored for use in decoding. Details on the spatial domain disparity compensation can be found in Section V.

Unlike previous photo collection compression schemes, we evaluate and generate the predictive difference between each pair of compensated reference block and the target one in the frequency domain. Rather than the decoded pixel values of input JPEG images, in this step we use the entropy decoded DCT coefficients from the input JPEG image as the target information. We also transform each compensated reference block to the DCT domain followed by the scalar quantization. The resulting quantized DCT coefficients are subtracted from the target ones. The generated residues are coded by the context adaptive arithmetic coding method. Finally, the coded residues and parameters are mixed up to generate the coded binary file. Since all operations generating the target files are invertible, lossless recovery of the original JPEG files is guaranteed. We will further present details of the frequency domain redundancy reduction in Section VI.

Fig. 3 shows the corresponding decoding process. After parsing the prediction structure, the intra-coded root image in the MST is first decoded. For each block, quantized DCT coefficients are recovered by adding decoded residues to the DCT transformed and quantized intra-compensated predictions. They are then inversely quantized and DCT transformed, resulting in recovered pixels of the block which are buffered as reference for subsequent decoding. For each inter-coded image, quantized DCT coefficients are also recovered by adding decoded residues to the compensated signal in the frequency domain where the compensated signal is generated by global and local compensations. After the inverse quantization and DCT, we get the pixels of the original JPEG coded image. The JPEG binary file of the image, on the other hand, is recovered by re-compressing the quantized DCT coefficients using the entropy coding method in JPEG.

Note that there is a clustering process before the presented scheme in Fig. 2 when dealing with relatively large scale

image sets. In this case, we will first cluster a set into small collections via a K-means based clustering method similar to [26] in which the distance between two images are defined as the average distance of matched SIFT descriptors. Then for each small collection, our presented scheme is applied. Though our MST-based prediction determination is also able to perform clustering, this will be very time consuming.

In the following, three modules in our hybrid lossless compression scheme, feature-domain determination of the prediction structure, spatial-domain disparity compensation, and frequency-domain redundancy reduction, will be introduced in greater detail.

#### IV. FEATURE-DOMAIN DETERMINATION OF PREDICTION STRUCTURE

Unlike natural video sequences which have strong temporal correlations, images in a collection usually have loose correlations and may vary in rotation, scale, and illumination. The inter-image disparities in image collections can be more complicated than those in videos. Traditional pixel-level disparity measurements, *e.g.* MSE, are not capable of effectively measuring the correlation between images.

Similar to [26], we introduce the feature-domain similarity to measure the inter-image correlation by the distance of their SIFT descriptors [40] to deal with large geometric transformations and luminance changes. A SIFT descriptor describes the distinctive invariant feature of a local image region, which consists of the location, scale, orientation, and feature vector. The key-point location and scale are determined by finding the maxima and minima of the difference of Gaussian filtered signals. The feature vector is a 128-dimensional vector which characterizes the local region by the histogram of the gradient directions, and the orientation denotes the dominant direction of the gradient histogram. SIFT descriptors have been demonstrated to have a high level of distinctiveness and thus are widely used in image search and object recognition.

We approximate the prediction cost between images by the average distance of their SIFT descriptors. Taking two images

$I_r$  and  $I_t$  as an example, the prediction cost  $p_{r,t}$  by using  $I_r$  to predict  $I_t$  is determined by the average distance of the set of matched SIFT descriptors  $S_{r,t}$  between these two images:

$$p_{(r,t)} = \frac{1}{|S_{r,t}|} \sum_{\forall (s_r^k, s_t^k) \in S_{r,t}} D(s_r^k, s_t^k) \quad (1)$$

where  $s_r^k$  and  $s_t^k$  are SIFT descriptors detected from  $I_r$  and  $I_t$ , respectively,  $(s_r^k, s_t^k)$  denotes the  $k^{th}$  pair of matched SIFT descriptors in the set  $S_{r,t}$ , and  $|A|$  denotes the number of elements in the set  $A$ .  $D(s_r^k, s_t^k)$  denotes the distance between the pair of descriptors,

$$D(s_r^k, s_t^k) = \frac{1}{|\mathbf{v}_t^k|} \|\mathbf{v}_r^k - \mathbf{v}_t^k\|^2 \quad (2)$$

where  $\mathbf{v}_r^k$  and  $\mathbf{v}_t^k$  denote the 128-dimensional feature vectors of  $s_r^k$  and  $s_t^k$ , respectively.

Let  $\mathbb{I} = I_1, I_2, \dots, I_N$  denote an image set which contains  $N$  correlated images. We estimate the predictive cost between each pair of images using equation (1) and use the directed graph to model the problem of prediction structure [21]. Fig. 4 shows one example. In this figure, each node denotes one image and the arrows (directed edges) denote the prediction paths between images. We first generate the directed graph as shown in Fig. 4 (a) and assign each arrow a predictive cost according to Equation (1). Then an MST can be deduced from the directed graph by minimizing the total predictive cost [41], as illustrated in Fig. 4 (b). At last, the prediction structure of the images in  $\mathbb{I}$  is determined by depth-first traversing the MST, as denoted in Fig. 4

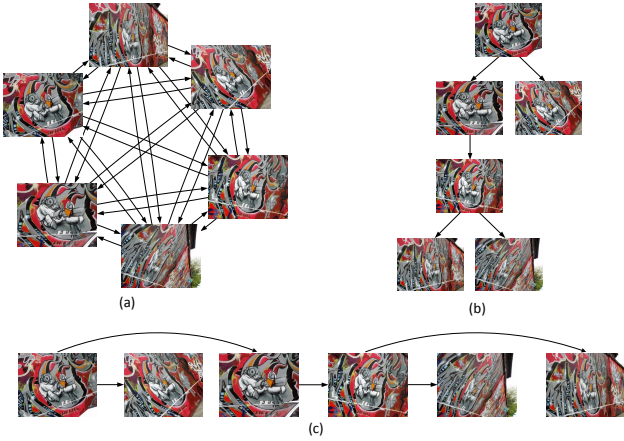


Fig. 4 Feature-based determination of prediction structure. (a) Directed graph in which each node denotes one image and arrows denote the prediction paths between images. (b) The MST deduced from the directed graph (a) by minimizing the total predictive cost. (c) Prediction structure determined by depth-first traversing the MST.

The prediction structure includes the prediction order of each image and the index of its reference, which needs  $N + (N - 1) = 2N - 1$  numbers ( $N$  is the number of images) to present. We code the numbers using 8-bit integers without loss.

## V. SPATIAL-DOMAIN DISPARITY COMPENSATION

Given the prediction structure of an image set  $\mathbb{I}$ , we then perform the spatial-domain disparity compensation to better exploit the correlations between images as well as image regions. As illustrated in Fig. 2, the root image in each MST is intra coded and compensated in the local compensation module. The other images are coded in reference to their parent images. Both the global and local compensations are performed on the reference image to approximate the corresponding target image.

Let  $I_t$  and  $I_r$  denote a pair of target and reference images, respectively. We thus formulate the disparity compensation in the spatial domain as

$$I_{r,t} = \begin{cases} \mathbf{L}_{r,t}(I_t), & \text{if } I_t \text{ is root,} \\ \mathbf{L}_{r,t}(\mathbf{G}_{r,t}(I_r)), & \text{otherwise.} \end{cases} \quad (3)$$

where  $\mathbf{G}_{r,t}$  denotes the global compensation,  $\mathbf{L}_{r,t}$  is the local compensation used to further reduce the small disparity between the target and the globally compensated reference images, and  $I_{r,t}$  is the compensated image when using image  $I_r$  to predict image  $I_t$ .

### A. Global Compensation

Inspired by [25], the global compensation in our scheme consists of two transformations, the geometric  $\mathbf{H}_{r,t}$  and photometric  $\mathbf{P}_{r,t}$  transformations, to deal with global geometric and illumination disparities, respectively.

$$\hat{I}_{r,t} = \mathbf{G}_{r,t}(I_r) = \mathbf{P}_{r,t}(\mathbf{H}_{r,t}(I_r)) \quad (4)$$

Let  $\bar{I}_{r,t}$  denote the deformed image of  $I_r$  after the geometric transformation,  $\bar{I}_{r,t} = \mathbf{H}_{r,t}(I_r)$ . For each pixel  $\bar{I}_{r,t}(x, y)$  in image  $\bar{I}_{r,t}$ , we compute the corresponding coordinates  $(x', y')$  in image  $I_r$  as

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad (5)$$

$$(x', y') = (u/w, v/w), \quad (6)$$

where the  $3 \times 3$  matrix is a homography,  $(u, v, w)$  and  $(x', y', 1)$  are the homogeneous coordinates after and before the transformation, and  $(x', y')$  are the final coordinates. Since the transformed coordinates  $(x', y')$  can be fractional, further interpolation is performed to get each warped pixel

$$\bar{I}_{r,t}(x, y) = \mathbf{F}(I_{r,t}, x', y'), \quad (7)$$

where  $\mathbf{F}$  is a linear interpolation algorithm.

The homography matrix is deduced from the matched SIFT feature points between images  $I_r$  and  $I_t$  using the RANSAC method [42]. We thus obtain image  $\bar{I}_{r,t}$  by the geometric transformation  $\mathbf{H}_{r,t}$  of image  $I_r$ .

We further introduce the photometric transformation  $\mathbf{P}_{r,t}$  if the images in the collection have large illumination variations. For each pixel  $\bar{I}_{r,t}(x, y)$  in  $\bar{I}_{r,t}$ , we adopt the linear transformation in [25] to generate the corresponding transformed pixel  $\hat{I}_{r,t}(x, y)$  in image  $\hat{I}_{r,t}$  as

$$\hat{I}_{r,t}(x, y) = a\bar{I}_{r,t}(x, y) + b, \quad (8)$$



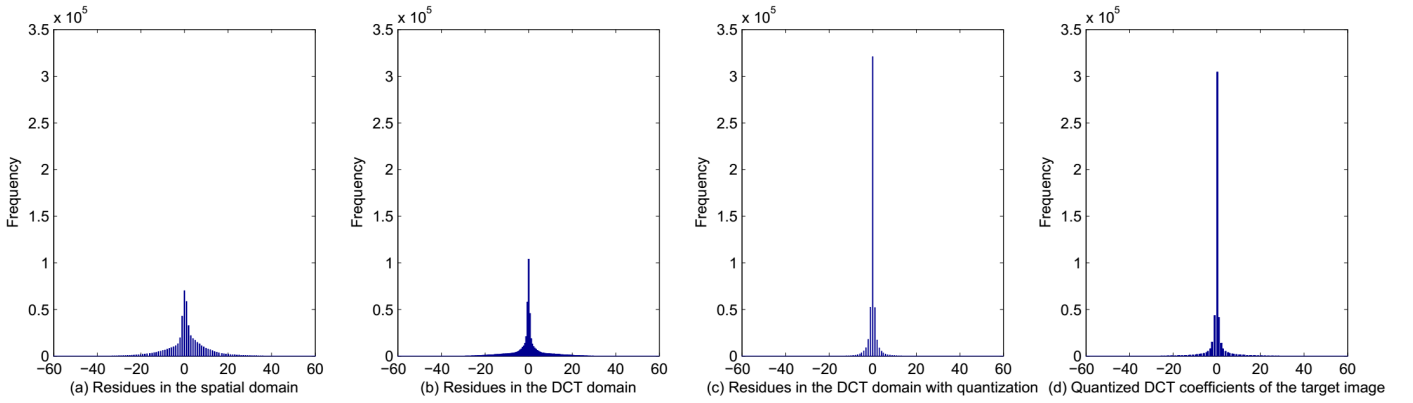


Fig. 5 Distributions of coefficients and residues. Distributions of (a) residues generated in the spatial domain, (b) residues generated in the frequency domain, (c) residues generated in the frequency domain with quantization, and (d) quantized DCT coefficients of the original JPEG coded image, respectively.

where  $a$  and  $b$  are the scale and offset parameters, respectively. These parameters are determined by minimizing the total prediction errors

$$\arg \min_{a,b} \sum_{(x,y) \in M} \|I_t(x,y) - (a\bar{I}_{r,t}(x,y) + b)\|^2, \quad (9)$$

where  $I_t(x,y)$  and  $\bar{I}_{r,t}(x,y)$  are the luminance values of the target image and the geometrically transformed image at the rounded positions  $M$  of the matched SIFT features. We solve (9) via linear regression in which  $a$  and  $b$  are estimated using least squares approximation.

In our global compensation, there are a total of ten parameters, eight from  $\mathbf{H}_{r,t}$  and two from  $\mathbf{P}_{r,t}$ . Since these parameters are fractional numbers, different floating-point implementations may lead to slightly different reconstruction values. For compatibility, we convert these parameters into integers by multiplying  $2^{32}$  and rounding. The resulting numbers are directly stored in 64-bit integers. At the decoder side, these parameters are normalized by  $2^{32}$  before the linear transformations. Thus, we can generate the same globally compensated image as the encoder with fixed precision parameters.

### B. Local Compensation

As shown in Fig. 2, the local compensation consists of two modules, the inter and intra compensations, which follows the method adopted in the HEVC coding standard [9].

We introduce inter compensation to further reduce the remaining small displacements between the target image  $I_t$  and the globally compensated one  $\hat{I}_{r,t}$ . We perform the block-based motion compensation with adaptive sizes  $s \in S = \{64 \times 64, 64 \times 32, 32 \times 64, 32 \times 32, 32 \times 16, 16 \times 32, 16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8\}$ . Intra compensation is performed for regions which cannot be well aligned by the global compensation or do not have content in the reference image  $\hat{I}_{r,t}$ . We adopt 35 modes (DC, Planar and 33 directional modes) in HEVC intra prediction to adaptively generate the reference block with minimal prediction error from neighboring pixels of previous coded blocks in the target image  $I_t$ . Please refer to [43] for detailed information on the 35 modes in HEVC intra

prediction. For each  $8 \times 8$  block, the best prediction mode, either inter or intra, is selected in the mode decision model to minimize the rate cost. The rate cost here is determined by the coding bits of motion vectors, modes, and residues, where the residues are calculated in the frequency domain as introduced in the following section. Note that the modes and motion vectors are coded in the same way as HEVC.

### VI. FREQUENCY-DOMAIN REDUNDANCY REDUCTION

After the disparity compensation, the redundancy between the target and compensated blocks will be reduced by calculating the residual signal. In all previous image set compression schemes, the residual signal is generated in the spatial domain. However, since the quantization step is skipped in the lossless compression, residues generated in the spatial domain usually lead to a heavily-tailed distribution as illustrated in Fig. 5 (a) and consequently make the following entropy coding less efficient in comparison with the lossy coded coefficients of the JPEG-coded image (as illustrated in Fig. 5 (d)).

On the other hand, JPEG coded images are quantized during lossy compression. It enables us to introduce the quantization in the frequency-domain redundancy reduction [11] [44]. In our scheme, we generate the residual signal between quantized DCT coefficients of the target and compensated blocks. As shown in Fig. 2, the quantized DCT coefficients of the target JPEG coded image can be acquired by the JPEG entropy decoding.

We then perform the same  $8 \times 8$  DCT transform as well as the quantization of the JPEG coded image on the corresponding compensated block. The residual signal is generated by calculating the difference between the two sets of quantized DCT coefficients. Fig. 5 (c) shows the distribution of the residual signal generated in the frequency domain. Compared with the heavy tailed distribution of the residues generated in the spatial domain as shown in Fig. 5 (a), the one generated in the frequency domain is significantly light tailed and thus greatly helps enhance the efficiency of the subsequent entropy coding.

Though the quantization module is introduced, we ensure lossless recovery of the original JPEG coded images as the

quantization is performed on the compensated signal instead of the residual one. All the prediction steps for lossless compression of the JPEG coded images as shown in Fig. 3 are invertible.

## VII. ENTROPY CODING

In our scheme, the residual signal of each block is coded by the arithmetic encoder. Rather than directly utilizing an existing entropy coding method, we design new context adaptive statistical models with regard to the statistical distribution of the quantized DCT residues, since the distribution of the quantized DCT residues (as denoted in Fig. 5 (c)) are different from those of signals generated in lossless (as illustrated in Fig. 5 (b)) or lossy (as shown in Fig. 5 (d)) image coding .

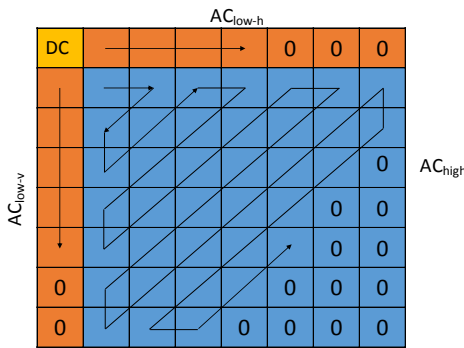


Fig. 6 Four partitions of an  $8 \times 8$  DCT block,  $DC$ ,  $AC_{low-h}$ ,  $AC_{low-v}$ , and  $AC_{high}$ , respectively.

In our arithmetic coding, we focus on exploiting the intra-block redundancy since most of the inter-image and inter-block redundancies are reduced in the disparity compensation. As shown in Fig. 6, we divide the 64 coefficients in each  $8 \times 8$  DCT block into four classes,  $DC$ ,  $AC_{low-h}$ ,  $AC_{low-v}$ , and  $AC_{high}$ , containing DC coefficient, AC coefficients in the first column, AC coefficients in the first row, and the remaining 49 AC coefficients, respectively. We also introduce three counters,  $N_{low-h}$ ,  $N_{low-v}$ , and  $N_{high}$ , denoting the numbers of non-zero (NNZ) coefficients in  $AC_{low-h}$ ,  $AC_{low-v}$ , and  $AC_{high}$ , respectively.

After statically partitioning the 64 coefficients according to their positions, we further adaptively categorize the symbols to be coded according to their causal neighbors (*i.e.*, contexts). Let  $S$  be a symbol to be coded, we define  $M_S^0$  as the probability distribution of  $S$ ,  $M_S^1$  as the probability distribution of  $S$  when a single context is involved, and  $M_S^2$  as the probability distribution of  $S$  when two contexts are considered. Given the probability distribution  $M_S^n$ , we can encode  $S$  using the arithmetic coding denoted as  $ArithCode(S, M_S^n)$ , where  $n = \{0, 1, 2\}$ .

Algorithm 1 summarizes the process of our arithmetic coding for  $N_{high}$  and  $AC_{high}$ . For each block, we first use  $M_{N_{high}}^1$  for  $N_{high}$  in which the average of the  $N_{high}$  values of the neighboring blocks above and to the left (if available) are used as the context. Then  $N_{high}$  is coded with arithmetic coding

using the probability distribution of  $M_{N_{high}}^1$ . Coefficients of  $AC_{high}$  are coded in the zig-zag order. The trailing zeros of the scanned sequence are skipped using the counter  $N_{high}$ . Each time a non-zero coefficient is coded,  $N_{high}$  is subtracted by one until it equals zero. The non-zero coefficient of  $AC_{high}$  is represented as  $c = \{s, a, l\}$ , where  $s$  denotes the sign of  $c$ ,  $a$  denotes the absolute value of  $c$ , and  $l$  is the length of the binary code of  $a$ .  $s$  is encoded with  $M_s^0$ .  $l$  is encoded with  $M_l^1$  with  $N_{high}$  as the context. The binary value of  $a$  is coded in bit plane. Let  $b$  denote the  $k^{th}$  bit plane of  $a$ .  $b$  is encoded with  $M_b^0$ . Since the first bit of  $a$  is always 1, the index of bit plane to be coded ranges from  $l - 2$  to 0.

Similarly, coefficients in  $AC_{low-h}$  and  $AC_{low-v}$  are coded in the horizontal and vertical orders using  $N_{low-h}$  and  $N_{low-v}$  to skip the trailing zeros, respectively. Let  $(x, y)$  denote the coordinates of the last non-zero coefficient in  $AC_{high}$  in the scanning order,  $N_{low-h}$  is coded via  $M_{N_{low-h}}^2$  with the contexts  $N_{high}$  and  $x$ , and  $N_{low-v}$  is coded via  $M_{N_{low-v}}^2$  given the contexts  $N_{high}$  and  $y$ . DC and the non-zero coefficients of  $AC_{low-h}$  and  $AC_{low-v}$  are coded the same way as  $AC_{high}$ .

### Algorithm 1 Encoding of $N_{high}$ and $AC_{high}$

```

1:  $ArithCode(N_{high}, M_{N_{high}}^1)$ 
2:  $n \leftarrow N_{high}$ 
3: for each  $c \in AC_{high}$  and  $n > 0$  do
4:    $\{s, a, l\} \leftarrow c$ 
5:   if  $c \neq 0$  then
6:      $ArithCode(s, M_s^0)$ 
7:      $ArithCode(l, M_l^1)$ 
8:     for  $k = (l - 2) \rightarrow 0$  do
9:        $b \leftarrow \{the\ k^{th}\ bit\ of\ a\}$ 
10:       $ArithCode(b, M_b^0)$ 
11:   end for
12:    $n \leftarrow n - 1$ 
13: else
14:    $ArithCode(0, M_l^1)$ 
15: end if
16: end for
```

## VIII. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our hybrid lossless compression scheme for JPEG coded image sets. Ten image collections as shown in Fig. 7 with different kinds of disparities are used in the following tests. Collection A contains photos of a natural scene with limited panning and rotation, Collection B contains photos with zooming in and out, Collection C contains two people in a slow motion, Collection D is a set of indoor images with different sleep posture, Collection E is composed of multi-view images<sup>1</sup>, images in Collection F have large rotation changes<sup>2</sup>, Collection G contains images with different illuminations<sup>2</sup>, Collection H is a set of frames from a video sequence in fast motion<sup>3</sup>, and Collection I and J are large scale outdoor image sets<sup>4</sup>.

<sup>1</sup>Available at [45].

<sup>2</sup>Available at <http://www.robots.ox.ac.uk/%7Evvgg/data/data-aff.html>.

<sup>3</sup>First 8 frames from HEVC test sequence "BasketballDrillText".

<sup>4</sup>Available at <http://evg.ethz.ch/research/symmetries-in-sfm>.

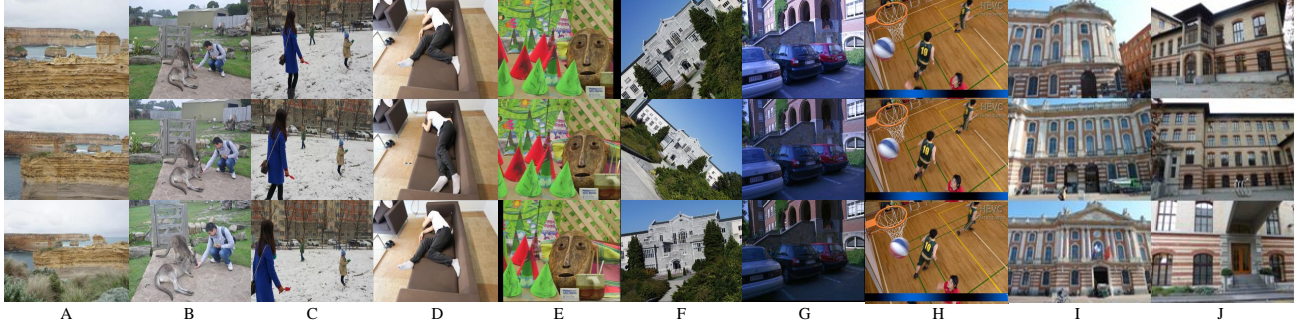


Fig. 7 Sample images of 10 test image collections.

Table I summarizes the properties of each collection, *e.g.* resolution and format. In our test, we convert Collection E to H to JPEG format using the library in the Independent JPEG Group(IJG) [46]. The quality factor of all the JPEG coded files is 90.

TABLE I Properties of the test photo collections.

Collection	A	B	C	D	E
Number	9	8	17	10	9
Resolution	960x640	4288x2848	2048x1152	720x480	1800x1500
Orig Format	JPEG	JPEG	JPEG	JPEG	PPM
Collection	F	G	H	I	J
Number	13	6	8	99	250
Resolution	800x640	900x600	832x480	multiple	multiple
Orig Format	PPM	PNG	YUV	JPEG	JPEG

Number: number of photos in each collection.

Orig Format: original format of photos in each collection.

The coding performance is measured by the bit-saving of the compressed bitstream (with a size of  $S_c$ ) compared to the size of the set of the original baseline JPEG files  $S_o$ .

$$bit-saving = (1 - \frac{S_c}{S_o}) \times 100\%. \quad (10)$$

To the best of our knowledge, we are the first to propose a lossless re-compression scheme for JPEG coded photo collections, so there is no related image set compression scheme for us to compare with. Existing lossless image set compression schemes that were not specifically designed for JPEG coded photos will not reduce the bit rate of JPEG coded collection but usually resulting in a larger file size [38]. In the following tests, we first demonstrate the overall performance of our scheme in comparison with several state-of-the-art individual JPEG recompression methods. We then evaluate the efficiency of the feature-based prediction structure, disparity compensation, and the entropy coding in our scheme. The impact of JPEG quality factor and image resolution is discussed later. Lastly, we analyze the complexity of our scheme .

#### A. Overall Performance

We evaluate the overall performance of our scheme with respect to several state-of-the-art lossless JPEG recompression

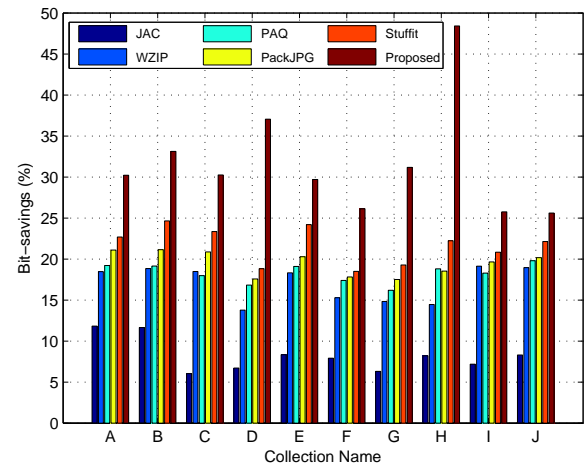


Fig. 8 Bit-savings of our scheme in comparison with some individual JPEG recompression methods.

methods, including JPEG arithmetic coding (JAC), WinZip<sup>5</sup>, PAQ<sup>6</sup>, PackJPG<sup>7</sup> and StuffIt<sup>8</sup>. Fig. 8 shows the compression performance in terms of bit-savings compared with the file sizes of the original JPEG coded photo collections. Among the six compression schemes, our proposed one achieves the best performance by exploring both inter and intra image correlations. The bit-saving of our scheme ranges from 25.6% to 48.4%. The average bit-saving of our scheme is 31.74%, which outperforms JAC, WinZip, PAQ, PackJPG, and StuffIt by 23.49%, 14.69%, 13.47%, 12.28% and 10.08%, respectively.

We also show the performance of JPEG2000 [4], H.264 [8] and H.265 [9] lossless coding profiles (individual compression of each image). Compared with the file size of the original JPEG coded photo collections in Table II, all these three schemes increase the file sizes.

<sup>5</sup>WinZip 19.0 Pro(11294) 32 bit.

Available at <http://www.winzip.com/win/en/downwz.html>

<sup>6</sup>Available at <https://cs.fit.edu/~mmahoney/compression/paq808.zip>

<sup>7</sup>PackJPG v2.5j Available at <http://www.elektronik.htw-aalen.de/packjpg/>

<sup>8</sup>Stuffit DELUXE 2010 for Windows, version 14.0.1.27. Available at <http://my.smithmicro.com/stuffit-deluxe-windows.html>



TABLE II Bit-savings (%) of JPEG2000, H.264 and H.265 (lossless mode) in comparison with the file sizes of the original JPEG coded images.

Collection	A	B	C	D	E
JPEG2000	-126.17	-134.02	-131.19	-142.33	-147.99
H.264	-131.09	-159.47	-136.15	-145.48	-176.11
H.265	-168.42	-197.50	-168.49	-165.82	-207.52
Collection	F	G	H	I	J
JPEG2000	-112.18	-135.51	-135.69	-123.08	-110.14
H.264	-126.21	-143.60	-139.23	-122.15	-118.77
H.265	-139.78	-173.42	-148.94	-148.54	-137.45

### B. Efficiency of the Feature-based Prediction Structure

In this subsection, we evaluate the efficiency of the feature-domain determination of the prediction structure in our lossless compression scheme. Fig. 9 shows the coding performance with and without the feature domain prediction structure (FPS) using Collection F, respectively.

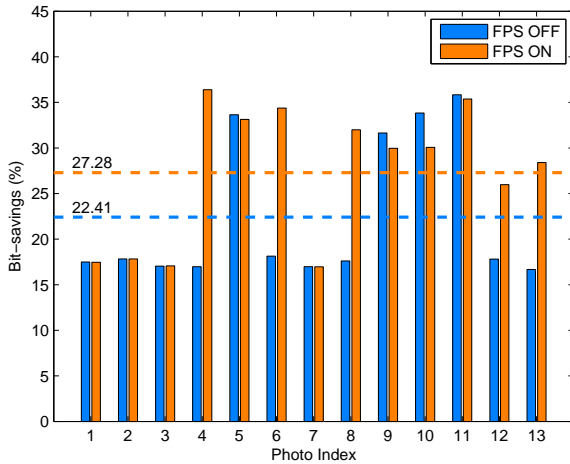


Fig. 9 Bit-savings with (“FPS ON”) and without (“FPS OFF”) feature-domain determination of prediction structure in comparison with the file size of the original JPEG coded collection F. The dashed lines show average bit-savings.

In this test, we first code the images in Collection F in the order of time stamp, resulting in the bit-saving of each image denoted as “FPS OFF”. The bit-savings of all frames with FPS are shown as “FPS ON”. Since different prediction orders are used, the bit-saving of each image fluctuates in different ways, resulting in the decrease (e.g. image 9) or increase (e.g. image 4) of bit-savings at different images for scheme “FPS ON”. The average bit-savings of these two schemes are presented by the dashed-lines in Fig. 9. In this case, the feature-domain prediction structure contributes 4.87% bit-saving on average to our lossless collection compression.

### C. Efficiency of the Disparity Compensation

In this subsection, we discuss the effectiveness of the spatial domain disparity compensation. The two stages, global and

TABLE III Bit-savings (%) in comparison with the file sizes of the original JPEG files with and without the global compensation.

Collection	A	B	C	D	E
GC ON	30.22	33.12	30.25	37.06	29.69
GC OFF	20.84	30.66	29.78	37.06	28.70
Delta	9.38	2.45	0.47	0	0.99
Collection	F	G	H	I	J
GC ON	26.13	31.18	48.41	25.75	25.61
GC OFF	17.40	28.13	48.41	-	-
Delta	8.73	3.05	0	-	-

GC ON (OFF): enable (disable) global compensation.

local compensation, are evaluated separately in the following tests.

We evaluate the efficiency of global compensation (GC) by enabling and disabling the GC in our scheme, respectively. Table III shows the comparison results for all the ten photo collections. We observe that the contribution of the GC varies for different collections with different levels of global disparities. The biggest improvements are observed for Sets A and F. This is because the global compensation reduces large disparities (zooming and rotation) between images, which are hard to compensate for by the block based motion estimation. On the other hand, our scheme does not benefit from GC when there is no global disparity between images such as in Collection D and H. Collection G has little geometric disparities but still achieves 3.05% improvement, owing to the photometric transformation. Note that Collection I and J have no results for “GC OFF”, due to the images with different resolutions can’t be aligned without GC in our scheme.

TABLE IV Bit-savings (%) in comparison with the file sizes of the original JPEG files with and without the local compensation.

Collection	A	B	C	D	E
LC ON	30.22	33.12	30.25	37.06	29.69
LC OFF	10.05	5.04	5.85	21.45	5.38
Delta	20.16	28.07	24.40	15.60	24.31
Collection	F	G	H	I	J
LC ON	26.13	31.18	48.41	25.75	25.61
LC OFF	4.44	14.09	27.05	8.24	7.96
Delta	21.69	17.09	21.36	17.52	17.65

LC ON (OFF): enable (disable) local compensation.

After global compensation, there are still small disparities between images. Thus block-based local compensation is adopted to reduce these small disparities. Table IV shows the efficiency of the local inter and intra compensations which reduce most redundancies and achieve 15.60% to 28.07% bit-savings.

### D. Efficiency of the Entropy Coding

We further evaluate the efficiency of our context adaptive entropy coder with respect to the entropy coding methods in JPEG and HEVC. We generate the test results by encoding our prediction residues of the quantized DCT coefficients using

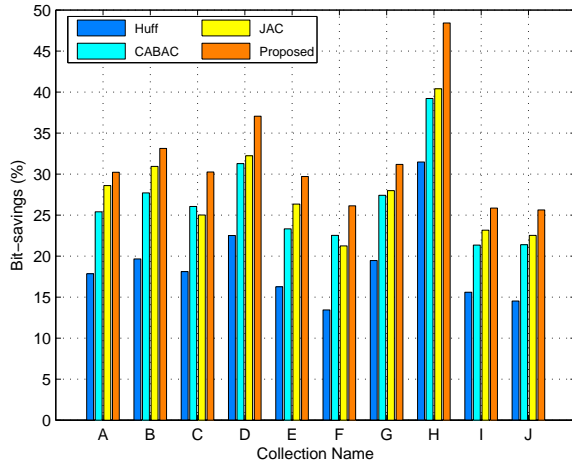


Fig. 10 Comparison of bit-savings with respect to the size of original JPEG coded files. Here “Huff”, “CABAC”, “JAC”, and “Proposed” denote the results generated using the Huffman-based coding method in the JPEG baseline, the CABAC based coding method in HEVC, the arithmetic coding method in JPEG extension and our proposed entropy coding method, respectively.

the Huffman coding method in JPEG baseline<sup>9</sup>, the CABAC<sup>10</sup> (content adaptive binary arithmetic coding) in HEVC, and the proposed algorithm.

Fig. 10 shows the comparison results in terms of bit-savings with respect to the size of the original JPEG coded files. Here “Huff”, “CABAC”, “JAC”, and “Proposed” denote the results generated using the Huffman-based coding method in the JPEG baseline, the CABAC based coding method in HEVC, the arithmetic coding method in JPEG extension, and our proposed entropy coding method, respectively. The results show that the proposed entropy coding method achieves the best performance among all four methods. “Huff” achieves the lowest bit-savings which mainly come from the previous feature and spatial domain modules as the input JPEG files are also coded by the Huffman coding method. Using “CABAC” for residue coding brings 5.8% – 9.1% bit-savings over “Huff” owing to the higher efficiency entropy coder with much more complicated context models. Due to the use of adaptive context models, “JAC” achieves higher bit-savings than “CABAC” in six out of the ten collections. Our proposed entropy coding method outperforms the other three entropy coding methods which further reduces up to 8.0% bits than “JAC” by exploring correlations between high and low frequency subbands.

#### E. Impact of Quality and Resolution of JPEG Coded Images

In the above tests, the quality factors of all the input JPEG coded images are 90. In this subsection, we discuss the impact of the quality of JPEG coded images on the performance of our scheme. We use Collection E to H to generate JPEG coded test image sets with quality factors ranging from 75 to 90. Fig.

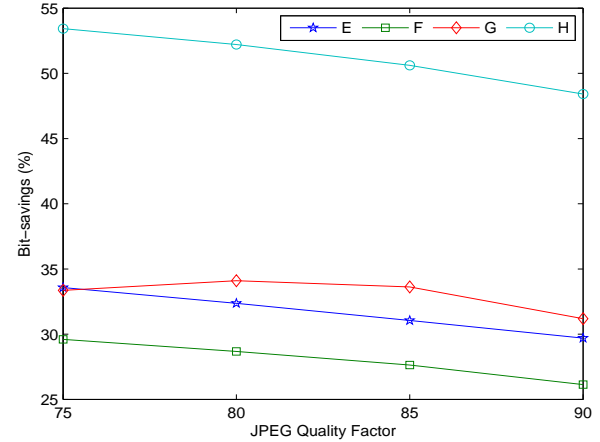


Fig. 11 Impact of quality factors of JPEG coded images (Collection E to H) on the coding performance of our scheme.

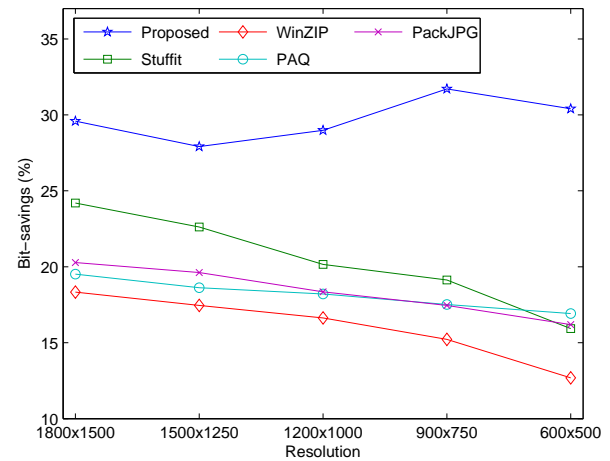


Fig. 12 Impact of resolution of JPEG coded images (Collection E) on the coding performance of different schemes.

11 shows the coding performance verses quality factor of our scheme. The results show that the bit-saving of our scheme gradually decreases with the increase of the quality factor. The reason for the decrease may be the higher the quality factors the more details in images, which may reduce the efficiency of prediction in the lossless coding scenario. In addition,

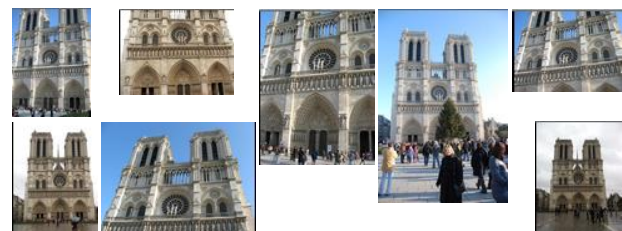


Fig. 13 Sample Images of Collection NotreDame.

<sup>9</sup>Implemented with the IJG library.

<sup>10</sup>Implemented with the HEVC test model (HM 12) [47].

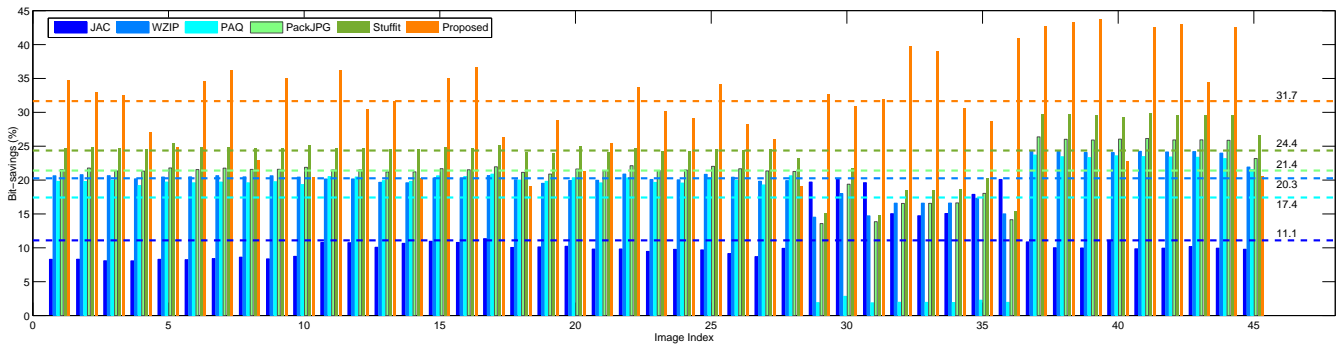


Fig. 14 Bit-savings of the 45 photos in Collection NotreDame. The dashed lines show the average bit-savings.

quantization errors are typically assumed to be independent from small distortions between current coefficients to be coded and reference ones, which may be another possible reason for the reduction of coding efficiency as the quality factor increases. We will to further study this phenomenon in the future work.

We also evaluate the impact of image resolution on the lossless recompression. We resize the PPM images in Collection E to different resolutions and then convert them to JPEG format with quality factor 90. Fig. 12 shows the coding performance of our scheme with respect to those of StuffIt, WinZip, PAQ and PackJPG. It shows that the bit-saving of our scheme fluctuates over a small range from 27.9% to 31.7% with the change of image resolution whereas the other four methods present a trend of monotone decrease with the reduction of image resolution.

We further demonstrate the robustness of our scheme in a very challenging case. We test the performance of our scheme using an image collection NotreDame (Fig. 13) which contains 45 photos downloaded from Flickr<sup>11</sup> taken by different people at 7 quality factors (75, 80, 82, 85, 90, 92, 96) and 8 resolutions ( $2592 \times 3872$ ,  $1122 \times 1496$ ,  $1704 \times 2272$ ,  $2448 \times 3264$ ,  $2304 \times 3072$ ,  $3072 \times 2304$ ,  $2272 \times 1704$ ,  $1944 \times 2592$ ). Thanks to the feature-based geometric compensation, our scheme naturally supports the transformation between images at different resolutions, resulting in a global compensated image at the same resolution as the target one, regardless of the size of the reference image. Fig. 14 shows the test result in bits savings for every photo. Among the six methods, our scheme performs the best and achieves 31.7% bit-savings on average.

#### F. Complexity

Our coding scheme is asymmetrical and the complexity of our encoder is much higher than that of the decoder.

On the decoder side, the computational cost comes from three parts, entropy decoding, local compensation, and global compensation. The complexity of the entropy coding as well as the local compensation is similar to that in HEVC. The complexity of global compensation is  $O(n)$ , where  $n$  is the number of pixels in an image. As shown in Fig. 15, the decoding complexity of our scheme is mainly determined by that of entropy decoding.

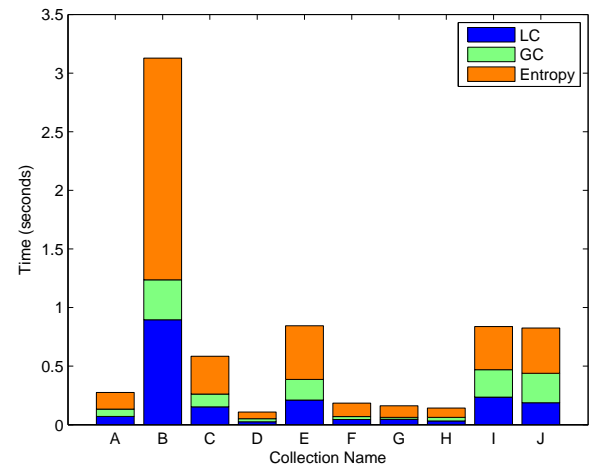


Fig. 15 Average computing time of each module for decoding a photo in each collection. “LC”, “GC” and “Entropy” denote the three modules of the decoder, *i.e.* local compensation, global compensation, and entropy decoding, respectively.

On the other hand, the encoding process is relatively time consuming as we adopt the intra and inter prediction methods in HEVC encoder, which has much higher complexity than the decoder and other individual JPEG compression methods due to the high computational cost of the joint optimization of partition and mode decision in local compensation. As shown in Fig. 16, our encoder spends 56.2% to 82.6% of the encoding time in the local compensation. Note that Collection J spends much more time on MST generation due to the  $O(n^2)$  complexity.

We would like to point out that our encoding process can be greatly accelerated via parallel computing. Due to the lossless nature, we can perform disparity compensation between each pair of reference and target images simultaneously so that the computational time of the disparity compensation of one image set can be reduced to that of a single image. Furthermore, the block-based intra block compensation as well as inter block compensation can also be processed in parallel to greatly speed-up the encoding process.

<sup>11</sup>Available at <http://www.flickr.com/>

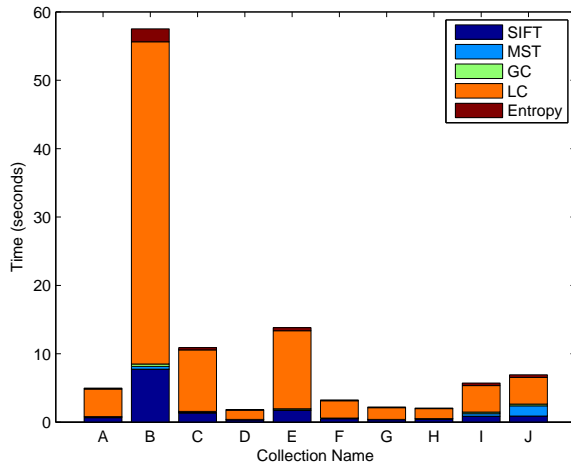


Fig. 16 Average computing time of each module for encoding a photo in each collection. “SIFT”, “MST”, “GC”, “LC” and “Entropy” represent SIFT extraction, MST-based prediction structure determination, global compensation, local compensation, and entropy encoding, respectively.

### G. Insertion, Deletion, and Random Access

Our scheme is able to support insertion, deletion, and the random access of an image set in different ways. For insertion and deletion, the most straightforward but also most time consuming way is to decode the whole image set and then perform the desired operation by re-encoding the updated set. On the other hand, there are alternative ways to support these operations. Taking insertion as an example, we can insert an image by encoding the new image individually or encoding it using the decoded root image as the reference. To delete an image, we can delete it directly if it is a leaf node or decode its parent node as well as its children nodes and then re-encode the children nodes only using the parent node as the reference. In terms of the MST updating, when edges need to be deleted from or inserted to the weighted directed graph, we can update the MST via the fast algorithms [48], [49] without traversing the whole directed graph. For random access, we can control the access delay by restricting the depth of the MST. The deeper the tree is, the larger the decoding delay but the higher the coding efficiency. The trade-off between the coding efficiency and operation speed can be determined according to different application requirements. We would like to point out the idea the depth-limited MST presented in both [26] and [50] which provides a good trade-off between the decoding delay and bit reduction can be used to control the difficulty in dealing with these operations. The fewer layers there are in a MST, the more easily the operations perform.

## IX. CONCLUSION

In this paper, we propose a new hybrid compression method to further reduce the file size of JPEG coded image collections without any fidelity loss. In our proposed method, we determine the prediction structure of each image collection by a feature domain distance measure. The disparity between

images is then reduced by joint global and local compensations in the spatial domain. In the frequency domain, the redundancy between the compensated and target images is reduced and the remaining weak intra correlations are further exploited in our entropy coding. By exploiting the correlations in the feature, spatial, and frequency domains, our scheme achieves up to 48.4% bit-savings and outperforms all state-of-the-art JPEG recompression schemes. We believe it can greatly reduce the storage cost for backup and archive of JPEG coded image collections for both personal and cloud applications.

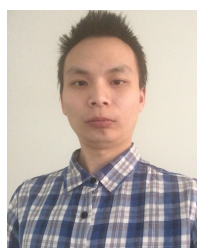
In this paper, we focus on the efficient compression method for a set of clustered JPEG images. We notice that the clustering can be time consuming if one collection is too large. Possible solutions may involve advanced fast clustering methods and introducing assistant information. For example, we can make use of the time stamps or GPS information in the meta data of images to separate a large collection to smaller ones. We would like to pay attention to reduce the complexity of the clustering module for large scale image sets in our future work. Besides, the performance of our proposed scheme could be further improved in several ways. First, we could speed up the encoding and decoding process by introducing parallel techniques. Second, we can further reduce the complexity of the local compensation in our scheme by not only leveraging some fast algorithms proposed for HEVC but also reducing complexity by direct operating on the JPEG coded DCT coefficients. Finally, we notice that the feature-based distance approximation may not be always efficient. In the future, we would like to investigate advanced distance metrics in which the number as well as the overlapped area of matched features are taken into account. We may also introduce a light weight version of the distance metrics so that the pixel-domain distance between two images can be measured much more accurately at low computational cost.

## REFERENCES

- [1] [Online]. Available: <http://jennstrends.com/instagram-statistics-for-2014/>
- [2] [Online]. Available: <http://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9>
- [3] G. K. Wallace, “The JPEG still picture compression standard,” *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. xviii–xxiv, 1992.
- [4] C. Christopoulos, A. Skodras, and T. Ebrahimi, “The JPEG2000 still image coding system: an overview,” *Consumer Electronics, IEEE Transactions on*, vol. 46, no. 4, pp. 1103–1127, 2000.
- [5] F. Dufaux, G. J. Sullivan, and T. Ebrahimi, “The JPEG XR image coding standard,” *Signal Processing Magazine, IEEE*, vol. 26, no. 6, pp. 195–199, 2009.
- [6] M. J. Weinberger, G. Seroussi, and G. Sapiro, “The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS,” *Image Processing, IEEE Transactions on*, vol. 9, no. 8, pp. 1309–1324, 2000.
- [7] X. Wu and N. Memon, “Context-based, adaptive, lossless image coding,” *Communications, IEEE Transactions on*, vol. 45, no. 4, pp. 437–444, 1997.
- [8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H. 264/AVC video coding standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.
- [9] G. J. Sullivan, J.-r. Ohm, W.-j. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [10] A. Said, “Introduction to arithmetic coding-theory and practice,” *Hewlett Packard Laboratories Report*, 2004.



- [11] I. Matsuda, Y. Nomoto, K. Wakabayashi, and S. Itoh, "Lossless re-encoding of JPEG images using block-adaptive intra prediction," in *Signal Processing Conference, 2008 16th European*, Aug 2008, pp. 1–5.
- [12] N. Ponomarenko, K. Egiazarian, V. Lukin, and J. Astola, "Additional lossless compression of JPEG images," in *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, 2005, pp. 117–120.
- [13] I. Bauermann and E. Steinbach, "Further lossless compression of JPEG Images," in *Proc. Picture Coding Symposium (PCS 2004)*, 2004.
- [14] M. Stirner and G. Seelmann, "Improved redundancy reduction for JPEG files," in *Proc. Picture Coding Symposium (PCS 2007)*, 2007.
- [15] G. Lakhani, "Modifying JPEG binary arithmetic codec for exploiting inter/intra-block and DCT coefficient sign redundancies," *Image Processing, IEEE Transactions on*, vol. 22, no. 4, pp. 1326–1339, 2013.
- [16] D. Salomon and G. Motta, *Handbook of data compression*, 5th ed. Springer, 2009, ch. 5.15 PAQ, pp. 314–319.
- [17] [Online]. Available: <http://kb.winzip.com/kb/entry/72/>
- [18] [Online]. Available: <http://www.elektronik.htw-aalen.de/packjpg/>
- [19] [Online]. Available: <http://my.smithmicro.com/win/index.htm>
- [20] A. Schmieder, H. Cheng, and X. Li, "A study of clustering algorithms and validity for lossy image set compression," in *Proc. The 2009 International conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV'09)*, 2009, pp. 501–506.
- [21] C.-P. Chen, C.-S. Chen, K.-L. Chung, H.-I. Lu, and G. Y. Tang, "Image set compression through minimal-cost prediction structures," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, 2004, pp. 1289–1292 Vol.2.
- [22] Y. Lu, T.-T. Wong, and P.-A. Heng, "Digital photo similarity analysis in frequency domain and photo album compression," in *Proc. The 3rd International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2004, pp. 237–244.
- [23] O. Au, S. Li, R. Zou, W. Dai, and L. Sun, "Digital photo album compression based on global motion compensation and intra/inter prediction," in *Audio, Language and Image Processing (ICALIP), 2012 International Conference on*, 2012, pp. 84–90.
- [24] R. Zou, O. C. Au, G. Zhou, W. Dai, W. Hu, and P. Wan, "Personal photo album compression and management," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, 2013, pp. 1428–1431.
- [25] Z. Shi, X. Sun, and F. Wu, "Feature-based image set compression," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, 2013, pp. 1–6.
- [26] —, "Photo album compression for cloud storage using local features," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 4, no. 1, pp. 17–28, 2014.
- [27] S. Milani and P. Zanuttigh, "Compression of photo collections using geometrical information," in *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, June 2015, pp. 1–6.
- [28] S. Milani, "Compression of multiple user photo galleries," *Image and Vision Computing*, 2016.
- [29] Y. S. Musatenko and V. N. Kurashov, "Correlated image set compression system based on new fast efficient algorithm of Karhunen-Loeve transform," in *Proc. SPIE*, vol. 3527, 1998, pp. 518–529.
- [30] K. Karadimitriou and J. M. Tyler, "The centroid method for compressing sets of similar images," *Pattern Recognition Letters*, vol. 19, no. 7, pp. 585–593, 1998.
- [31] —, "Min-max compression methods for medical image databases," *ACM SIGMOD Record*, vol. 26, no. 1, pp. 47–52, 1997.
- [32] C.-H. Yeung, O. C. Au, K. Tang, Z. Yu, E. Luo, Y. Wu, and S. F. Tu, "Compressing similar image sets using low frequency template," in *Multimedia and Expo (ICME), 2011 IEEE international Conference on*, 2011, pp. 1–6.
- [33] T. Shao, D. Liu, and H. Li, "Inter-picture prediction based on 3d point cloud model," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3412–3416.
- [34] D. Perra and J. Frahm, "Cloud-scale image compression through content deduplication," in *BMVC*, 2014.
- [35] X. Song, X. Peng, J. Xu, G. Shi, and F. Wu, "Cloud-based distributed image coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, no. 12, pp. 1926–1940, 2015.
- [36] H. Yue, X. Sun, J. Yang, and F. Wu, "Cloud-based image coding for mobile devicestoward thousands to one compression," *Multimedia, IEEE Transactions on*, vol. 15, no. 4, pp. 845–857, 2013.
- [37] H. Wang, M. Ma, and T. Tian, "Effectively compressing near-duplicate videos in a joint way," in *Multimedia and Expo (ICME), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1–6.
- [38] H. Wu, X. Sun, J. Yang, and F. Wu, "Lossless compression of JPEG coded photo albums," in *Visual Communications and Image Processing Conference, 2014 IEEE*, Dec 2014, pp. 538–541.
- [39] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still image data compression standard*. Springer Science & Business Media, 1993.
- [40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.
- [41] Y. J. Chu and T. H. Liu, "On the shortest arborescence of a directed graph," *Science Sinica*, vol. 14, pp. 1396–1400, 1965.
- [42] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [43] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, "Intra coding of the HEVC standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1792–1801, 2012.
- [44] M. Karczewicz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 637–644, July 2003.
- [45] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1, June 2003, pp. 1–195.
- [46] The independent jpeg group. [Online]. Available: <http://ijg.org>
- [47] S.-i. S. Ken McCann, Benjamin Bross and W.-J. Han, "Encoder-Side Description of HEVC Test Model (HM)," in *JITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-C402*, Oct. 2010.
- [48] G. Pollatos, O. A. Telelis, and V. Zissimopoulos, "Updating directed minimum cost spanning trees," in *Experimental Algorithms*. Springer, 2006, pp. 291–302.
- [49] M. R. Henzinger and V. King, "Maintaining minimum spanning forests in dynamic graphs," *SIAM Journal on Computing*, vol. 31, no. 2, pp. 364–374, 2001.
- [50] Y. Ling, O. C. Au, R. Zou, J. Pang, H. Yang, and A. Zheng, "Photo album compression by leveraging temporal-spatial correlations and hevc," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 1917–1920.



**Hao Wu** received the B.S. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is currently working toward the Ph.D. degree in electrical engineering in Tianjin University, Tianjin, China. His research interests include image/video compression and computer vision.



**Xiaoyan Sun** (M'04-SM'10) received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1997, 1999, and 2003, respectively. Since 2004, she has been with Microsoft Research Asia, Beijing, China, where she is currently a Lead Researcher with Internet Media Group. She has authored or co-authored more than 60 journal and conference papers and ten proposals to standards. Her current research interests include image and video compression, image processing, computer vision, and cloud computing. Dr. Sun was a recipient of the Best Paper Award of the IEEE Transactions on Circuits and Systems for Video Technology in 2009.





**Jingyu Yang** (M'10) received the B.E. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2003, and the Ph.D. (Hons.) degree from Tsinghua University, Beijing, in 2009. He has been a Faculty Member with Tianjin University, Tianjin, China, since 2009, where he is currently an Associate Professor with the School of Electronic Information Engineering. He was with Microsoft Research Asia (MSRA), Beijing, in 2011, within the MSRAs Young Scholar Supporting Program, and the Signal Processing Laboratory, cole Polytechnique Fdrale de Lausanne (EPFL), Lausanne, Switzerland, in 2012, and from 2014 to 2015. His current research interests include image/video processing, 3-D imaging, and computer vision.

Dr. Yang was selected into the program for New Century Excellent Talents in University from the Ministry of Education, China, in 2011, the Elite Peiyang Scholar Program and Reserved Peiyang Scholar Program from Tianjin University, in 2012 and 2014, respectively, and the Promotion Program for Innovative Talents from Tianjin Science and Technology Commission in 2015.



**Wenjun Zeng** (M'97-SM'03-F'12) is a Principal Research Manager at Microsoft Research Asia, while on leave from Univ. of Missouri (MU) as a Full Professor. He had worked for PacketVideo Corp., Sharp Labs of America, Bell Labs, and Panasonic Technology prior to joining MU in 2003. He has contributed significantly to the development of international standards (ISO MPEG, JPEG2000, and OMA). He received his B.E., M.S., and Ph.D. degrees from Tsinghua Univ., the Univ. of Notre Dame, and Princeton Univ., respectively. His current research interest includes mobile-cloud video acquisition and analysis, social network/media analysis, multimedia communications/networking, and content/network security.

He is/was an Associate Editor of IEEE Trans. on Circuits & Systems for Video Technology (TCSVT), IEEE Multimedia (currently an Associate EiC), IEEE Trans. on Info. Forensics & Security, and IEEE Trans. on Multimedia (TMM), and is/was on the Steering Committee of IEEE Trans. on Mobile Computing (current) and IEEE TMM (2009-2012). He served as the Steering Committee Chair of IEEE Inter. Conf. Multimedia and Expo (ICME) in 2010 and 2011, and has served as the TPC Chair/co-Chair of several IEEE conferences (e.g., ChinaSIP'15, WIFS'13, ICME'09, CCNC'07). He will be a general co-Chair of ICME2018. He is currently guest editing a TCSVT Special Issue on Visual Computing in the Cloud - Mobile Computing, and was a Guest Editor (GE) of ACM TOMCCAP Special Issue on ACM MM 2012 Best Papers, a GE of the Proceedings of the IEEE's Special Issue on Recent Advances in Distributed Multimedia Communications (January 2008) and the Lead GE of IEEE TMM's Special Issue on Streaming Media (April 2004). He is a Fellow of the IEEE.



**Feng Wu** (M'99-SM'06-F'13) received the B.S. degree in Electrical Engineering from XIDIAN University in 1992. He received the M.S. and Ph.D. degrees in Computer Science from Harbin Institute of Technology in 1996 and 1999, respectively. Now he is a professor in University of Science and Technology of China and the dean of School of Information Science and Technology. Before that, he was principle researcher and research manager with Microsoft Research Asia.

His research interests include image and video compression, media communication, and media analysis and synthesis. He has authored or co-authored over 200 high quality papers (including several dozens of IEEE Transaction papers) and top conference papers on MOBI-COM, SIGIR, CVPR and ACM MM. He has 77 granted US patents. His 15 techniques have been adopted into international video coding standards. As a co-author, he got the best paper award in IEEE T-CSVT 2009, PCM 2008 and SPIE VCIP 2007. Wu has been a Fellow of IEEE. He serves as an associate editor in IEEE Transactions on Circuits and System for Video Technology, IEEE Transactions on Multimedia and several other International journals. He got IEEE Circuits and Systems Society 2012 Best Associate Editor Award. He also serves as TPC chair in MMSP 2011, VCIP 2010 and PCM 2009, and Special sessions chair in ICME 2010 and ISCAS 2013.