Optimizing Compression and Storage of JPEG Images

R.K.Prakash[1,] G.Adhitya[2], M.Prashanth[3], KarthikeyaTripathi[4]
[1,2,3,4] UG students,
Department of Computer Science Engineering,
SRM University, Ramapuram, Chennai – 89
[1]rkprakashbw@gmail.com, [2]adithyaiyer@gmail.com
[3]Prashanthmuthuram@gmail.com,[4]kartiktripathi028@gmail.com

M.Prabu[5]
[5]*Assistant Professor,*
Department of Computer Science Engineering,
SRM University, Ramapuram, Chennai – 89
Mail.id- manavalanprabu@gmail.com

Abstract— The explosion in digital photography poses a significant challenge when it comes to photo storage for both personal devices and the Internet. In this paper, we propose a novel lossless compression method to further reduce the storage size of a set of JPEG coded correlated images. In this method, we propose jointly removing the inter-image redundancy in the feature, spatial, and frequency domains. For each album, we first organize the images into a pseudo video by minimizing the global predictive cost in the feature domain. We then introduce a disparity compensation method to enhance the spatial correlation between images. Finally, the redundancy between the compensated signal and the corresponding target image is adaptively reduced in the frequency domain. Moreover, our proposed scheme is able to losslessly recover not only raw images but also JPEG files. Experimental results demonstrate the efficiency of our proposed lossless compression, which achieves more than 12% bit-saving on average compared with JPEG coded albums.

Keywords—Image compression, lossless, JPEG, image set, album compression, image coding.

1. Introduction

The increasing number of digital photos on both personal devices and the Internet presents a significant challenge in terms of storage and bandwidth. The most popular way to reduce the size of a photo is JPEG compression [1]. Though some digital cameras support raw data storage of captured photos, the majority of captured photos are compressed using image compression schemes, e.g. JPEG and JPEG2000. Though feasible for random access, individual compression ignores the redundancy between images when dealing with a set of correlated images.

To reduce the inter-image redundancy between images, a number of image set compression schemes have been pro- posed. When a set of images are similar enough, a representative signal (RS) (e.g., an average image) can be generated from the images and the differential signals and the RS are coded [2]–[4]. For lossy compression, the images can also be organized into a pseudo video in which the prediction structure is determined via a MST (minimal spanning tree) and then coded like a video sequence [5]–[7] using video coding, e.g. HEVC [8]. Though efficient, all the schemes are presented to compress raw images rather than the JPEG coded ones. As most of the photos today are compressed and stored in JPEG format, a lossless recompression scheme for JPEG coded photo albums is desirable to reduce storage costs.

In this paper, we propose the first lossless compression method to further compress a set of JPEG coded correlated images. Our proposed scheme ensures identical reconstruction of both image content and JPEG binaries which is required for applications such as data backup, cloud storage. For each JPEG coded album, we propose a hybrid algorithm to remove the inter-image redundancy in the feature, spatial, and frequency domains jointly. We introduce the feature-based measurement to determine the prediction structure so that we can well handle the variations in rotation, zooming, and illumination. The disparity between images is then compensated by global and local alignments in the spatial domain. The final compressed file is achieved by reducing the redundancy in the frequency domain. Experimental results demonstrate the advantage of our scheme in terms of achieving much higher coding efficiency and lossless representation of the JPEG coded files.

The rest of this paper is organized as follows. Section II gives a brief overview of our proposed lossless compression scheme. Our feature-based prediction structure, spatial domain disparity compensation, and frequency domain redundancy removal are presented in Section III, Section IV, and Section V, respectively. The experimental results are presented in Section VI for performance evaluation. Section VII concludes this paper.

2. Proposed Method

Lossless compression of JPEG coded images is a tough topic. There are two key problems here: how to model the correlation among images and how to compress the differential signal without loss so as to achieve a reduced file size with regard to the JPEG coded o n e .

Fig. 1 illustrates the basic idea of our lossless compression scheme. In this scheme, we compress JPEG coded photo albums by make use of the redundancy between images inside each album. At the encoder side, for each input photo album, we first determine the prediction structure with regards to the feature level similarity between each pair of images. Based on the prediction structure, we then reduce the disparity between adjacent images by joint global and local compensation. The predictive difference between the compensated image and the target one is evaluated and compressed in the frequency domain. The generated frequency residues are coded via the entropy coding method used in JPEG to produce the coded album.
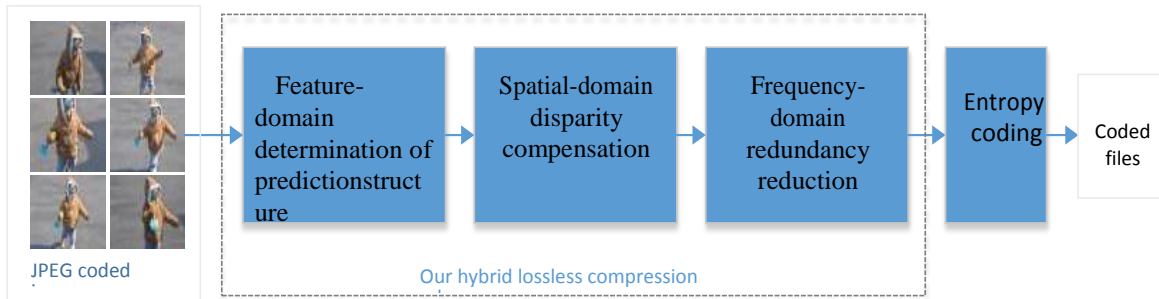
Fig. 1.Illumination of, hybrid lossless compression scheme (encoding process).

The corresponding decoding process is illustrated in Fig.2.Thepredictionorderisfirstdecodedfromthecodedbitstream. Each predictively compressed (i.e. inter-coded) image can be decompressed individually like JPEG coded images.

In the following section, the three modules in our hybrid lossless compression (as shown in Fig. 1), feature-domain determination of prediction structure, spatial-domain disparity compensation, and frequency-domain redundancy reduction, will be introduced in detail.
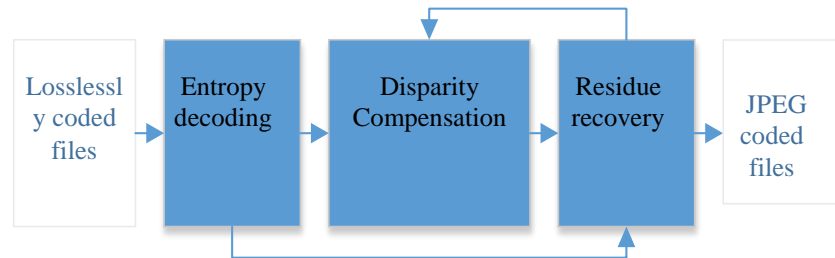

Fig. 2. Illumination of the decoding process of our proposed hybrid lossless compression scheme.

Feature domain determination of prediction structure

In contrast to natural video sequences, which have an inherent temporal correlation, images in an album have disordered and loose correlations. Moreover, the inter-image disparity in photo albums are more complicated than that in natural videos, as the images may vary a great deal in rotation, zoom, and illumination. The traditional pixel-level disparity measures,

e.g. MSE, are not capable enough to deal with photo albums. In our proposed scheme, we measure the similarity between images by the distance of their SIFT descriptors [9] to avoid the impact of large motion offsets and luminance changes, similar to [7]. A SIFT descriptor describes the distinctive invariant feature of a local image region, which consists of a

location, scale, orientation, and feature vector. The key-point location and scale are determined by finding the maxima and minima of difference of Gaussian signals. The feature vector is a 128-dimensional vector, which characterizes the local region by the histogram of 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 widely used in image search and object recognition.

We approximate the predictive cost between images by the distance of their SIFT descriptors. Taking two images, I_n and I_m, as an example, the predictive cost $p_{n,m}$, using I_n to predict I_m, is determined by the average distance of the set of matched SIFT descriptors $S_{n,m}$ between these two images.
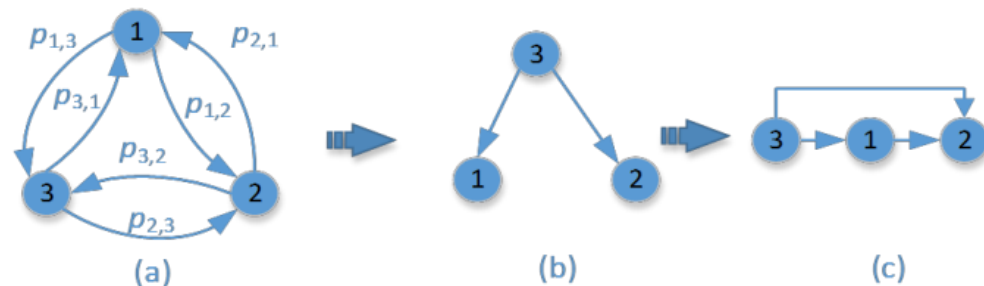

Fig. 3.   MST-based determination of prediction structure.

Let II = I_1, I_2, ..., I_N denote an image set that contains N correlated images. We estimate the predictive cost between each pair of images using 1 and generate the directed graph, as exemplified in Fig. 3 (a). Then an MST can be deduced from the directed graph by minimizing the total predictive cost [10], as shown in Fig. 3 (b). The prediction structure of the images in I is determined accordingly by depth-first traversing the MST, as denoted in Fig. 3 (c). The depth can be limited   to one when random access of a single image is required

Spatial-domain disparity compensation

Given the prediction structure of an images et II, we can reorganize the images into a pseudo video and directly use motion compensation and estimation technologies for   videocoding to reduce the disparity between references and target images. However, as mentioned before, the disparity in an image set can be more complicated than that in a video sequence. To solve this problem, we adopt the prediction method proposed in [7]to exploit the correlation between images. We generate the compensated image from I_n to I_m as

$$C_{n,m} = F_{n,m}(\boldsymbol{T}_{n,m} \times \boldsymbol{H}_{n,m} \times I_n) \qquad (3)$$

Where homograph transform $\boldsymbol{H}_{n,m}$ and photometric transform $\boldsymbol{T}_{n,m}$ are used to reduce the geometric and illumination dis- parities, respectively, and $F_{n,m}$ denotes a block-based motion compensation function. The deformation matrix $\boldsymbol{H}_{n,m}$ is deduced from the matched SIFT descriptors between images I_n and I_m using the RANSAC method[11]. The photometric

transformation matrix $T_{n,m}$ is deduced from the grey information of images I_n and I_m using all the pixel values between the images. With regards to the robustness of the matched SIFT descriptors; we use only the matched key-point values to estimate the matrix. More details on generating the two matrices can be found in [7].

We noticed that there are still smaller local disparities, e.g., local shifts, between the transformed image and image I_m, so we further employ a simple block-based motion estimation scheme to deal with local disparities, resulting in the final compensated image Cn,m.
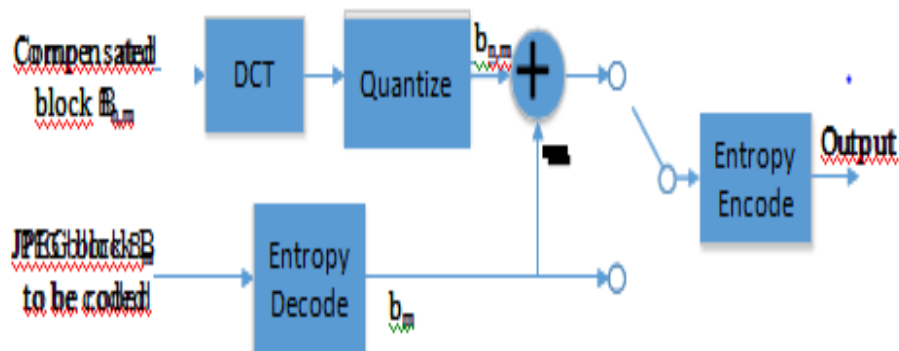

Fig. 4.Illumination of frequency-domain redundancy reduction (encoding process).

Frequency-domain redundancy reduction

With the compensated image $C_{n,m}$, a straight forward way to reduce the redundancy is calculating the difference between $C_{n,m}$ and I_m and coding the difference signal without loss, with or without transforms. However, compressing such difference signals without loss requires considerable bits and usually leads to a larger size of a bit stream in comparison with that of the original JPEG set. As demonstrated in Table I, using HEVC lossless to encode the difference signals generates a much larger file size than that of the original JPEG set.

In contrast to all the previous lossy or lossless image set compression schemes, we propose generating the difference signal in the frequency domain, as shown in Fig. 4. Our frequency-domain redundancy reduction is performed at the block-level. To align with JPEG compression, we fix the block size at 8 8. Let B_m denote a JPEG coded block. After entropy decoding, we get the block b_m containing the JPEG coded DCT coefficients without inverse quantization. Our frequency domain prediction tries to find a DCT block that is similar to b_m. We thus propose first performing a DCT transform to the compensated block $B_{n,m}$, and then preforming the JPEG quantization. The resulting quantized DCT coefficients $b_{n,m}$ are used to predict b_m and generate a difference signal at the DCT domain, which will be entropy

encoded later. Since the range of (*bm bn, m*) is smaller than that of *bm*, the coded   bits are also smaller than the original JPEG. Notice that though we introduce a quantization step to the compensated block, the original JPEG coded DCT block *bm* can be recovered without loss on the decoder side since the quantization is performed on the predicted signal instead of the coded   difference.

Additionally, the prediction structure, the DCT prediction modes of all the blocks, and the compensation parameters, including matrix H and T, and the motion vectors of inter-coded blocks, are all encoded and transmitted to the   decoder.

TABLE I
SIZES OF PHOTO ALBUMS (KB)

Album	A	B	C	D	E	F
Original JPEG	2061.73	756.37	826.02	1341.18	1769.33	4405.27
RAR	2028.77	756.87	826.38	1326.13	1765.45	4374.93
HEVC-Lossless	4532.89	2069.86	2146.46	3289.44	4418.64	11740.0
Our method	**1691.67**	**660.04**	**708.21**	**1202.70**	**1571.08**	**3905.72**

3.Experiment Results

To the best of our knowledge, there are no comparable methods designed for compressing a JPEG coded image set, not to mention that our scheme can recover not only the pixel information of images but also JPEG coded bits. For example, the HEVC lossless compression scheme in   Table I can exactly recover the pixel values but cannot recover the original JPEG binary files since it ignores the Meta data as well as the quantization information in the original JPEG files. Therefore, we make a comparison with the generalized lossless compression algorithm RAR (5.01 64bit) [12].

We evaluated the performance of our scheme using six photo albums (JPEG baseline, JFIF format, YUV420 down- sampled), as shown in Fig. 5. These photo albums were shot at different places (both indoor and outdoor) and targets (people, buildings, and scenes), with different levels of motion, zoom, and illumination. Among them, albums B and C were downloaded from [13] and [14], and album E is a personal album captured at the Summer Palace [7].

Table I shows the album sizes of the original JPEG coded photo album, RAR compression, HEVC lossless, and our proposed scheme. The RAR can save very limited bits but produces larger file sizes for coding albums B and C. HEVC lossless cannot even reduce

the set size on any of the albums. In contrast, our scheme constantly outperforms both JPEG and RAR. Compared with JPEG coded album, our scheme achieves 12.96% bit-saving on average. The bit-savings of RAR and our scheme can be found in Fig.   6.

The main complexity comes from calculating the prediction cost in the feature domain (40%) and motion estimation in the spatial domain (40%). The encoding time for each image in the test datasets is 4.29s on average without   optimizations.
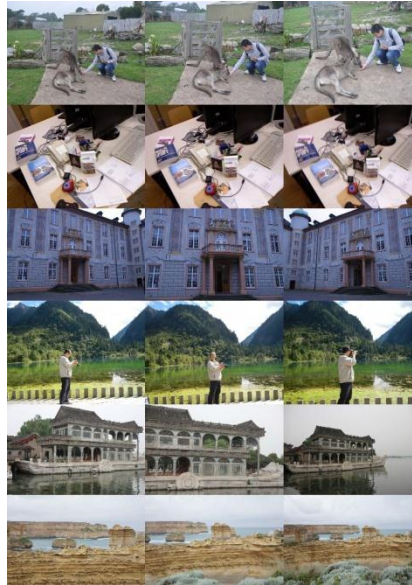

Fig. 5.Sample images of six tests photo albums. From top to bottom: A   (8 photos, 1024x768), B (11 photos, 640x480), C (8 photos, 768x512), D (6 photos, 1024x768), E (20 photos, 784x512), and F (22 photos,   1200x800).

4. Conclusion

In this paper, we have proposed the first lossless compression scheme for reducing the file size of JPEG coded photo albums. We propose a novel hybrid scheme to efficiently make use of the correlation between images in feature, spatial, and frequency domains, respectively. Compared with JPEG coded files, our scheme is able to achieve up to 17.95% bit-savings and meanwhile is capable of recovering both pixels and JPEG binaries without loss.

The efficiency of our proposed scheme can be improved in several ways. First, the homography and photometric trans- formation for global compensation is not robust enough for   a complicated scene and sudden change of viewpoints and illumination lighting. Also, we have only adopted a very simple motion estimation method (i.e. 8 8 block motion estimation without motion vector prediction). An advanced motion compensation scheme, e.g. the method in HEVC, can be used to enhance the compensation efficiency.
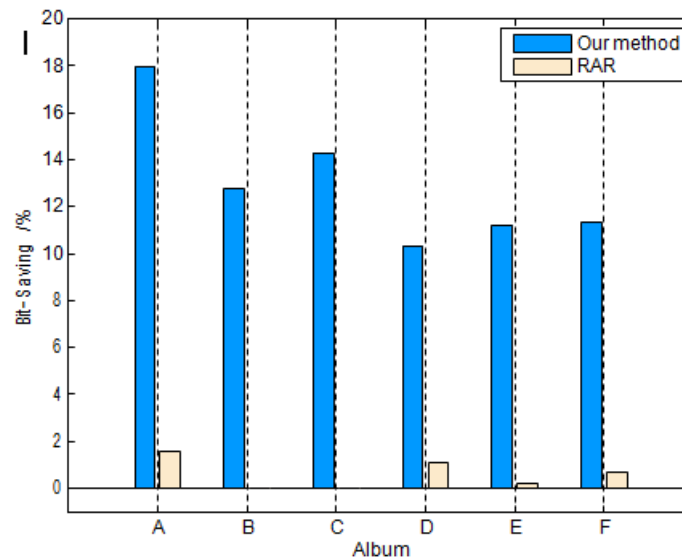
Fig. 6. Bit-savings of RAR and our scheme in comparison with the original JPEG coded photo albums

Similarly, better entropy coding methods can also be utilized. Finally, we would like to test our scheme on large-scale albums and extend our scheme to support big data compression.

References

[1] G. K. Wallace, "The JPEG still picture compression standard," Communications of the ACM, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[2] 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.

[3] 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. IEEE, 2011, pp.1–6.

[4] K. Karadimitriou and J. M. Tyler, "Min-max compression methods for medicalimagedatabases,"ACMSIGMODRecord,vol.26,no.1,pp.

[5] C.-P. Chen, C.-S.Chen, K.-L.Chung, H.-I. Lu, and G. Y.Tang, "Imageset compression through minimal-cost prediction structures." in ICIP,2004, pp.1289–1292.

[6] 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.   IEEE, 2013, pp.1428–1431.

[7] Z. Shi, X. Sun, and F. Wu, "Feature-based image set compression," in Multimedia and Expo (ICME), 2013 IEEE International Conference on. IEEE, 2013, pp.1–6.

[8] G. J. Sullivan, J.-r. Ohm, W.-j. Han, and T. Wiegand, "Overview of theHigh Efficiency Video Coding (HEVC) Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649– 1668, Dec. 2012.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International journal of computer vision, vol. 60, no. 2, pp. 91–110, 2004.

[10] Y. J. Chu and T. H. Liu, "On the shortest arborescence of a directed graph," Science Sinica, vol. 14, pp. 1396–1400, 1965.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigmfor model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol. 24, no. 6, pp. 381–395,1981.

[12] [Online]. Available: http://www.win-rar.com/download.html
[13] [Online]. Available: http://vision.in.tum.de/rgbd/dataset/freiburg1
[14] [Online]. Available: http://cvlab.epfl.ch/strecha/multiview