Image Set Modeling By Exploiting Temporal-Spatial Correlations and Photo Album Compression

Ruobing Zou*, Oscar C. Au, Guyue Zhou, Sijin Li and Lin Sun
*Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology, Hong Kong SAR, China
E-mail: {zouruobing, eeau, zhouguyue, sliae, lsunece}@ust.hk Tel: +852--23588529

Abstract — **with the advance of digital photographing technology, large amount of personal photos are created and stored online or in personal computers. To save storage space and transmission bandwidth, we proposed a new photo album compression scheme by using both intra prediction and inter prediction to reduce spatial and temporal redundancy. Specifically, we first cluster the images so that each cluster containing a set of similar images is a group of pictures (GOP) with variable length. A graph framework is proposed then, in which an optimal "IPPP…P" GOP structure is derived from every cluster by finding the minimum spanning tree (MST) at a minimum prediction cost. Finally, the photo album is compressed as a whole like a video sequence, by High Efficiency Video Coding (HEVC), according to the predictive order of the optimal structures. The experimental results show that our scheme achieves around 60% improvement over only using JPEG compression.**

I. Introduction

Today's digital photographing gains increasing popularity in people's daily life, especially by the Digital Single Lens Reflex (DSLR) cameras. Thanks to the fast developing digital camera technologies, costs of hardware keep declining - cameras of the same cost as years ago can make pictures with higher resolution and of course, bigger size. Nevertheless, brought about is a problem of storage volume expansion.

Bearing this in mind, we want to further compress the file size for photo albums. Recently, more attention is paid to the topic of personal digital image compression. Among these researches, Yang et al. developed a compact photo album compression method by first calculate the similarity distances of every two photos in frequency domain to perform photo album sorting and clustering so as to group similar photos together, and then the clustered photos are compressed by a MPEG-like algorithm. [1] Chia-ping et al. proposed a new scheme for compressing an object movie image (OM) set by building its minimal-cost prediction structure. [2] B. Gergel et al. also developed a graph-theoretic framework as a conclusion of most previous schemes. [3]

In this paper, a new method for digital album compression is proposed. We first divide a collection of digital photos into smaller sets by clustering, in which a graph is formed to represent the correlation between or within images. A minimum spanning tree is constructed and trimmed for each graph, so as to get an optimal encoding structure. We then encode the whole album with the order of these structures by the latest video coding technique HEVC, which has advantage in high definition video/image compression. Our method provides a practical and effective way to store and manage large amount of digital photos.

II. Photo Album Analysis

An analysis of the characteristics of digital photographing will help us understand the behavior of albums so as to establish the theoretical framework. A digital photo album is defined as one or several folders containing pictures exported from digital cameras. The images can either be: (1) uncompressed pictures like RAW files or NEF files (by NIKON), or (2) more often, compressed pictures like JPEG and TIFF files. This paper will focus on the latter ones. But in whatever format, personal photos generally have the following characteristics.

First, photo albums are supposed to consist of similar images, since photographers are used to taking a bunch of photos for a same scene or event. Redundancy thus needs to be reduced for compact compression. Corresponding to video coding technologies such as in the H.264/AVC and the HEVC, one effective method is intra prediction, which uses various spatial prediction modes to exploit spatial redundancy within a single picture. Images predictive in this way are referred to as intra images (I-images). The other is inter prediction, which uses motion vectors to reduce temporal redundancy between different frames to achieve much higher compression rates. Likewise, images are called inter image (P-image).

However, personal photos share some unique characteristics. Unlike video sequence which has a strict linear chronicle order with strong inter-frame correlation, a digital album usually contains a number of images that has variable inter-image correlation – namely, the similarity between two chronically adjacent images is unknown. Because between images are larger time intervals than that of a video sequence.

Additionally, unlike MRI/CT images, face images and OM images in [2] where images are largely similar, photographers change scenes frequently and objects always move. Scene recurrence is also common as cameras may turn back to previous scene. These unique characteristics make it inefficient to directly copy existing video compression techniques. Therefore we first pre-process the album by clustering and re-ordering to make it better for compression.

III. PROPOSED IMAGE SET MODELING SCHEME

A. Images Clustering

Digital album is first clustered in order to put similar images into a same group. Image clustering has been a heated research topic for which a number of methods were developed, especially for retrieval and indexing of images. Since this part is not the focus of our research, we adopted an established clustering method in [1]. Without having to decompress JPEG photos, the method calculates the energy histograms in frequency domain directly. Then the similarity distances of every two images are calculated and adaptive clustering algorithms is performed to group similar images together. By using this method, photo albums can be effectively clustered.

B. Graph Framework

For each cluster, let $S = \{I_1, I_2 \ldots I_n\}$ be a set of all n images in it; these images are of the same dimension and are in the same color space. We assume YUV color space in this paper. A weighted directed graph $G = \langle V, A\rangle$ is constructed from S, where the nodes $V = \{I_i \mid I_i \in S\}$ represent n images, and arcs $A = \{(I_i, I_j) \mid I_i, I_j \in V\}$ represent predictive coding: arc (I_i, I_j) directed from I_i to I_j means I_j is to be encoded/decoded and I_i is a reference image used to predict I_j. Associated with arc (I_i, I_j) is a weight $c(I_i, I_j)$, denoting the cost of prediction. Graph G has the following properties, as shown in an example graph G of 4 images in fig.1 (a):

1. Each node has a loop since i may be equal to j. In this case, arc (I_i, I_j) $(i=j)$ denotes intra prediction within a same picture.
2. For each arc (I_i, I_j) $(i \neq j)$, its corresponding inverted arc (I_j, I_i) also belongs to G, indicating two opposite inter prediction directions. Note that $c(I_i, I_j)$ and $c(I_j, I_i)$ may be different.
3. G has n nodes and $n(n-1)/2+n$ arcs.

Above is the graph model in a cluster. We want to find an optimal order of this image set, in which predictive coding algorithms can best reduce redundancy to achieve a low compression bit-rate and good picture quality. In other words, we seek for a sub-graph of G at minimum cost of iterating all nodes V.

The optimal structure should satisfy following conditions:

1. Given a cluster, an I-image serves as the root of the sub-graph which has only outgoing arc and no ingoing arc.
2. Located at the root should be the only loop in the sub-graph, representing the only I-image given an image set.
3. Other nodes each has one predecessor that will be used as a reference in prediction.

A MST by Chu/Liu and Edmond algorithm like in [2] would be an effective solution to find the sub-graph. But before that, graph G needs to be expanded to fit the algorithm.

First, a major difference between our condition and Chu/Liu and Edmond algorithm is that a loop's cost should also be counted into the total cost of the optimal structure, whereas Chu/Liu and Edmond algorithm discards all loops at the beginning. For this, we delete all loops and replace each of them by a mirror node and an arc. The mirror nodes $V_m = \{I_{mi} \mid I_{mi} = I_i\}$ respectively points to their corresponding real nodes, connected by arcs $A_m = \{(I_{mi}, I_i) \mid I_{mi} \in V_m, I_i \in V\}$. Each arc's cost is the same as the original loop's:

$$c(I_{mi}, I_i) = c(I_i, I_i) \quad (1)$$

representing the cost of intra prediction on the real node.

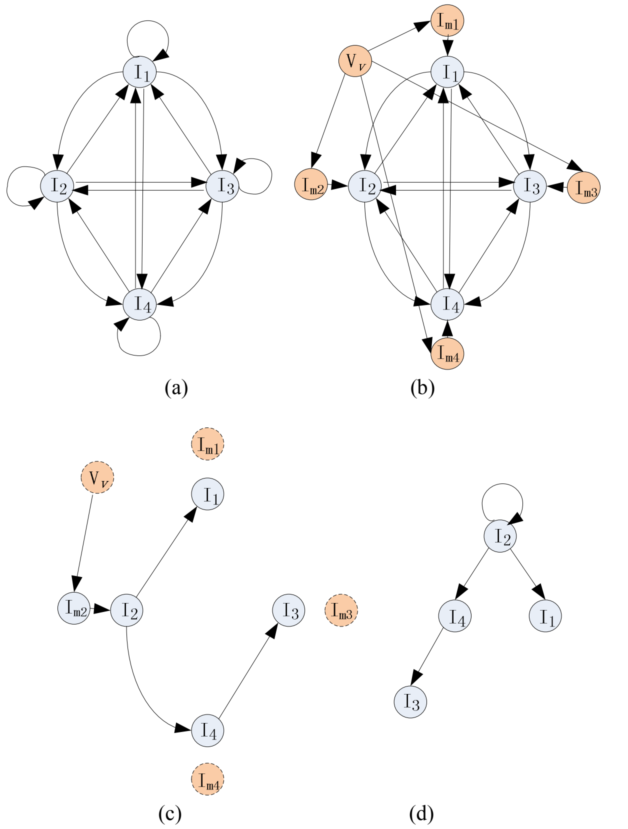


Fig.1 An example of the graph framework
(a) Graph G for a cluster; (b) The expanded graph G'; (c) The minimum spanning tree derived from G'; (d) Optimal structure.

Second, as we want to select a best root to start encoding with, a virtual node I_v is added to this graph as a global root. It only has out-going arcs $A_v = \{(I_v, I_{mi}) \mid I_{mi} \in V_m\}$ pointing to each mirror node with a cost equal to a value that is larger than the total cost of all other arcs:

$$c(I_v, I_{mk}) > \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} c(I_i, I_j) + \sum_{i=1}^{n} c(I_{mi}, I_i) \quad (2)$$

where $k = 1,2,\ldots n$. These two modifications make a new expanded graph G', as shown in the example in fig.1 (b). Its nodes V' and arcs A' are:

$$V' = V \cup V_m \cup \{I_v\}$$
$$A' = A \cup A_m \cup A_v \quad (3)$$

Then Chu/Liu and Edmond algorithm is applied to G' to build a MST as in fig.1(c). As we can see, the MST must be rooted at the virtual node I_v since it has only outgoing arcs. Also, there is one and only one successor of I_v, which is a mirror node I_{mk} $(I_{mk} \in V_m)$, otherwise any other path will cost less than an additional arc from I_v, valued larger than the total cost of any other path.

Finally, we delete I_v as well as the mirror node I_{mk}, which points to a real node I_k. This node I_k is the root of the optimal structure. Adding back the original loop onto I_k, we get the optimal coding structure of image set S, which is a MST with a loop, as in Fig.1 (d).

C. Cost Estimation and Complexity reduction

The cost $c(I_i, I_j)$ of an arc can be estimated in several ways. Ideally, encoding rate is a most close approximation, as it is the total bits spent on encoding picture I_j with I_i as a reference. However, this approach is time-consuming, and it doubles the work calculating $c(I_i, I_j)$ $(i \neq j)$ through motion estimation, motion compensation, transforming, and entropy encoding. These will be repeated in images compression later. A better estimation is performing motion estimation between I_i and its reference image I_j. Cost is defined as the sum of absolute differences (SAD). As for loops, $c(I_i, I_j)$ $(i = j)$ is set to be a picture's SAD by intra prediction.

Furthermore, we simplify this process to reduce complexity for motion estimation, as it takes up large proportion of arc cost computing time.

Considering the fact that every arc (I_i, I_j) has an invert arc (I_i, I_j), an already acquired motion vector from the motion estimation of (I_i, I_j) can be reused as a predictor of its invert arc. In detail, given a block b_j in I_j, a two-dimensional motion vector v is derived providing an offset from I_j to the best matching block b_i in a reference image I_i. Later as calculating $c(I_i, I_j)$, its reversed motion vectors v', pointing from I_i to I_j, is used as a motion vector predictor so that the motion estimation will start at the endpoint of v'. The rationale behind is that a motion vector points to the best matching block on the reference picture is likely to point to a similar block backward.

IV. PHOTO ALBUM COMPRESSION

With the help of the optimal encoding structure, we compress the digital photo album. Unlike images in medical applications that allow little quality degradation, online images or personal photos are tolerable to quality loss within some range. Therefore, lossy coding technique is taken in this paper, by discarding less important visual information to minimize file size while closely approximates the source files. [3] When we compress a clustered and ordered photo album, we treat it as a video sequence.

A. Review of the High Efficiency Video Coding (HEVC)

As most of the digital pictures are with high definition, we adopt HEVC as coding technique. Targeted at next-generation HDTV, HEVC standard is developed by the Joint Collaborative Team on Video Coding. It is expected to reduce bitrates by half than H.264/AVC with comparable image quality. Encoding algorithms may select between inter and intra coding for block-shaped regions of each picture. The prediction residual is then further compressed using a transform. Finally, the motion vectors or intra prediction modes are combined with the quantized transform coefficient information and encoded using arithmetic coding. [4]

Of course, the applicable video coding standard is not limited to HEVC, also applicable are the H.264/AVC, MPEG-4 etc.

B. Modified GOP Structure

In video coding, a group of pictures (GOP) specifies the order in which intra- and inter-frames are arranged. GOP encoding structure in HEVC starts with an intra coded I-frame, followed by P-frames forwardly predicted by its previous I-frame, or B-frames which can have forward and backward prediction by at most two references. [4]

In photo album compression, each cluster makes of an independent "IPPP…P" GOP. Each GOP begins with an I-image that is the root of the optimal structure; followed by P-images if any, including no B-image. Each image's reference has been specified beforehand according to the optimal structure described in the section III. A P-image is allowed to be predicted by either an I-picture or its precedent P-image. The encoding/decoding order is according to breadth-first search (BFS) order of the optimal structure. Since the size of each cluster is variable, we expect GOP to have variable length too.

Noted that the difference between our method and [2] lies in that they store the original image as the root, without compression, while in our scheme the root is intra coded by HEVC to further reduce the total encoding bit-rate. If a cluster contains only one picture, then it is encoded as an I-image and there is no P-image.

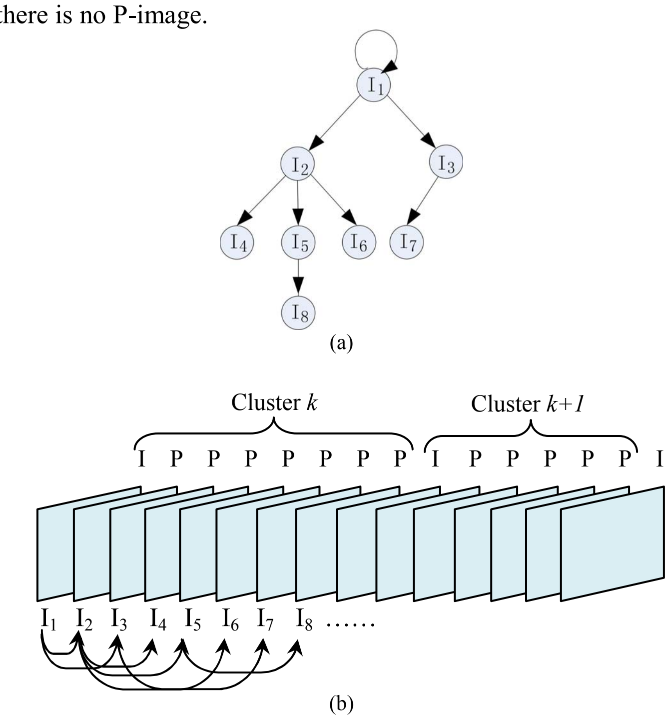


Fig.2 An example of the optimal structure and corresponding GOP. (a) An optimal structure marked in the BFS order (b) Part of the image sequence and $\{I_1, I_2, \ldots I_8\}$ is the corresponding GOP of the structure in (a).

V. EXPERIMENTAL RESULTS

In our experiment, we tested how our method works compressing JPEG pictures from digital albums stored in personal computers. The encoding/decoding software we use is HEVC Test Model HM 7.0.

Two photo albums are tested, each containing .JPG pictures of the same dimensions. The pictures are taken in daily life, including party events, portraits, scenery and architectures etc. Album #1 has 182 pictures taken by digital still camera. Album #2 has 71 pictures by DSLR. When quantization parameter (Qp) in HM 7.0 is set to be 32, the compression results of the two albums are in Table I. There is 59.8% improvement over the traditional JPEG format for photo album #1; and 53% improvement for photo album #2.

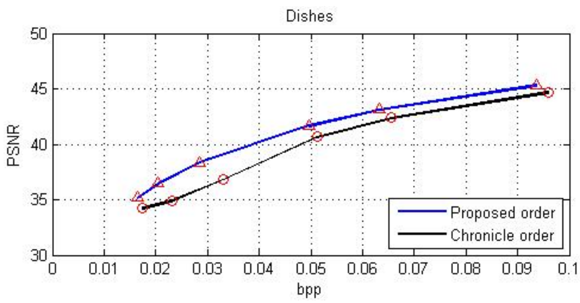
TABLE I
HEVC COMPRESSION RESULTS (IN BYTES)

Scheme	Traditional JPEG	MST+loop +HEVC	Improvement
Photo Album #1	256MB	103MB	59.8%
Photo Album #2	259MB	121MB	53.3%

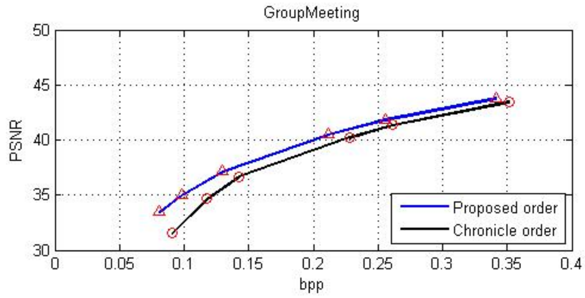
Another experiment is to compare the performance of the proposed order, i.e. a MST with a loop, with chronicle order. Chronicle order means predicting one by one according to picture taken time, as it is the default setting how pictures are stored in computers. We test two clusters separately in Fig. 3. Cluster "Dinner" contains 7 .JPG images taken by CANON EOS 60D, with the dimension of 1024x768. Cluster "Meeting" contains 29 .JPG images taken by NIKON D90, the dimension of which is 4288x2848. Theses clusters are not from two previous albums. The result shows that our method outperforms when images or in chronicle order. The results are shown in shown in Fig. 4 (a) and Fig. 4 (b).


(a)
(b)

Fig. 3 Clusters to be tested: (a)"Dishes" and (b)"Group meeting", more pictures in (b) are not shown above



(a)



(b)

Fig. 4 Performance in different predictive order

VI. CONCLUSIONS

In this paper, we proposed a new scheme to compress digital photo albums by exploiting both the inter image and the intra image correlations. The proposed method first clusters a photo album into several groups each consisting of similar images. Each cluster is modeled into a weighted directed graph, the arc weights of which are estimated as the predictive coding cost. The graph is then expanded to find a MST at minimum prediction cost. Having had the MST, we get the optimal encoding structure by trimming the MST and add a loop at the root. After that, the photo album is compressed by lossy coding technique the HEVC, where every cluster is considered as a GOP with "IPPP…P" structure, predicted in an order consistent with the optimal structure. As experimental results show, our method achieves significant improvement than the traditional photo storage technique, as well as gains than arranging in the chronicle order in terms of compression ratio.

ACKNOWLEDGMENT

This work has been supported in part by the Research Grants Council (RGC) of the Hong Kong Special Administrative Region, China. (GRF Project no. 610210)

REFERENCES

[1] Y. Lu, T.T. Wong, and P.A. Heng, "Digital photo similarity analysis in frequency domain and photo album compression," Mobile and Ubiquitous Multimedia; Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia, Vol. 83, pp. 237-244, 2004.

[2] C.P. Chen, C.S. Chen, K.L. Chung, H.I. Lu, and G. Tang. "Image set compression through minimal-cost prediction structures," in IEEE International Conference on Image Processing, pp. 1289-1292, 2004.

[3] B. Gergel, H. Cheng, C. Nielsen, and X. Li, "A unified framework for image set compression," in Proceedings of the 2006 International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada, USA, Volume 2, pp. 417-423, June 2006.

[4] B. Bross, W.J. Han, J. R. Ohm, G.J. Sullian, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 7," in JCT-VC 9th Meeting, Geneva, JCTVC-I1003_d2, 2012.

[5] C. Yeung, O. C. Au, K. T. Tang, Z. D. Yu, E. Luo, Y. Wu, and S.F. Tu, "Compressing similar image sets using low frequency template," in Proc. ICME, 2011, pp.1-6.