An Inter-Image Redundancy Measure for Image Set Compression

Xinfeng Zhang, Yabin Zhang, Weisi Lin
School of Computer Engineering
Nanyang Technological University, Singapore
Email:{xfzhang, wslin}@ntu.edu.sg

Siwei Ma, Wen Gao
School of Electronics Engineering and Computer Science
Peking University, Beijing, China
Email:{swma, wgao}@pku.edu.cn

***Abstract*—Image set coding improves the compression efficiency by reducing both intra- and inter-image redundancy. The key of success is to select representative image(s) to predict set of similar images. This paper proposes an inter-image redundancy measure for representative image selection in image set compression. In the proposed method, the inter-image redundancy is measured jointly by the extent of similar content (EOS) and the correlation of similar content (COS) shared in two images. We take the covered area of matched SIFT points to measure the EOS, and take the distance of the matched SIFT descriptors to measure the COS. The image with largest redundancy for the set is selected as the representative one to predict other images. Experimental results show that the proposed method can select better representative image, and achieve bitrate saving up to 9.2% and 20.8% compared with state-of-the-art image set compression method and HEVC inter coding method.**

Keywords—Image set compression; set redundancy; SIFT; cloud storage

I. INTRODUCTION

Image and video coding is important for most of the digital image/video applications, which achieves high compact expression for image/video by reducing various types of redundancy. Traditional image coding methods compress each image individually by intra prediction [1][2] and entropy coding to reduce spatial redundancy and statistical redundancy, respectively. For lossy compression, quantization technique is utilized to reduce the psychovisual redundancy, e.g., different quantization steps applied to coefficients in different frequency bands in JPEG compression [3]. In addtion, there are also inter-image redundancy among similar images (also named as set redundancy [4]), which is similar with that in videos [5][6]. With the development of cloud storage, a large amount of similar images are uploaded to the cloud but still compressed individually, and this consumes lots of storage space since common content shared by similar images are stored repeatedly. Therefore, an image set coding technique could further improve compression efficiency by jointly compressing a set of similar images to reduce the set redundancy among them.

During the last two decades, many image set compression methods have been proposed, which can be roughly classified into two categories according to the prediction structures, i.e., pairwise prediction method and sequential prediction method illustrated in Fig. 1. Pairwise prediction methods are usually with a star topological prediction structure, in which one representative image is selected or generated from one image set, compressed independently, and then the others (denoted as predicted images) in image set are compressed by referring to the reconstructed representative image. Karadimitriou et al. proposed some representative image generation methods, including max-min differential (MMD) method [4] and centroid method [7]. The MMD method generates two representative images, "min" image and "max" image, which are constructed by choosing the minimum and maximum pixel values for every pixel position, and then compresses difference between original pixel values and either the min or the max image (whichever is smaller) for every image in the set. The centroid method generates one representative image by averaging the pixel values in the same position among all the images, and then the average image and the difference images between the representative image and predicted images are compressed individually. Yeung et al. [8] take a low frequency template as the representative image by only averaging the low frequency components of all images to capture the common patterns in set of similar images while discarding variation content. However, these methods cannot handle images with large scale geometric deformations, and also need to compress extra representative image(s).

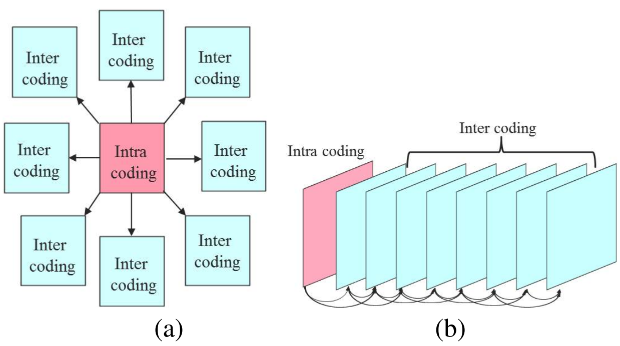


Fig. 1. Prediction structure for image set compression. (a) star topological prediction structure, (b) sequential prediction structure.

Based upon the success of video coding, Zou et al. proposed an image set compression and management method in [9], which arranges the set of similar images as a video sequence, and compresses them with HEVC inter coding directly. The coding efficiency for a given image set is mainly determined by the image coding order. Zou et al. take the sum of absolute difference (SAD) of motion compensation between any two images to measure their correlation, based on which the coding order is derived by finding a minimum spanning tree (MST). Due to irregular large scale motion existing among similar images which makes traditional motion estimation inefficient, Shi et al. [10] [11] take advantage of the distance of matched SIFT descriptors between any two images to measure their correlation. However, these methods increase the delay for image access severely, e.g., accessing the images in leaf node positions need decode all the dependent precedents. Therefore, the tree depth is usually constraint [11][12][13]. When the depth is one, the prediction structure is the same as the star topological prediction structure.

In this paper, we propose an inter-image redundancy measure to find the best representative image for image set compression. For an image, its inter-image redundancy relative to another image can be considered as how much similar content existing in it (denoted as the extent of similar content, EOS) and how correlated the similar content (denoted as the correlation of similar content, COS) being. We take the two aspects jointly to estimate the inter-image redundancy. For the EOS of an image, we take the size of the covered area by matched SIFT points on the other images to measure it. The larger the covered area in other images, the more its inter-image redundancy exists. For the COS of an image, we take the distance of matched SIFT descriptors between two images to measure it. The smaller distance of the matched SIFT descriptors is, the more correlated the covered area by SIFT points between the two images is. Based on the inter-image redundancy measure, we propose an image set compression method with star topological prediction structure by selecting the representative image with maximum inter-image redundancy, which is compressed with intra coding method. The other images are compressed by referring to the reconstructed representative images which is processed by geometric and photometric transform just as that in [11].

The remainder of the paper is organized as follows. In section II, we review the SIFT based image set compression framework. Section III introduces the proposed inter-image redundancy measure method and introduces the image set compression method in detail. Experimental results are reported in Section IV and Section V concludes the paper.

II. REVIEW OF SIFT-BASED IMAGE SET COMPRESSION

Scale Invariant Feature Transform (SIFT) [14] is an algorithm in computer vision to detect and describe local features in images, which is invariant to scale, rotation, illumination and viewpoint. Therefore, the SIFT descriptor is well suitable for finding the similar content among images by searching matched SIFT points. For an image $\mathcal{I}_n$, there are a set of SIFT points and the k-th SIFT descriptor is formulated as $f_n(k) = \{\mathbf{p}_n(k), \mathbf{v}_n(k)\}$, where $\mathbf{p}_n(k) = \{x_1, x_2, s, o\}$ represents the coordinate, scale and orientation parameters, while $\mathbf{v}_n(k)$ is a 128-dimensional local feature descriptor vector.

In [11], Shi et al. proposed an efficient image set compression framework based on SIFT descriptors. Firstly, the clustered similar images are organized into a minimum spanning tree (MST) based on the correlation of similar images, where the distance of matched SIFT descriptors is used to measure their correlation. Then, the set of similar images are arranged into a sequence according to the depth-first traversing of the MST. The root image is firstly intra coded without inter-image prediction. The other images are coded as inter ones predicted from their parents by geometric and photometric transform.

III. IMAGE SET COMPRESSION WITH SIFT-BASED INTER-IMAGE REDUNDANCY MEASURE

A. SIFT-based inter-image redundancy measure method

The prediction order plays an important role in image set compression for pairwise prediction structure and the sequential prediction structure. For an image set $S = \{\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_n\}$, the most straight-forward way to measure inter-image redundancy is to calculate the residual energy of the motion compensation between any two images as follows,

$$e(\mathcal{I}_i, \mathcal{I}_j) = \|\mathcal{I}_i - T(\mathcal{I}_j)\|^2 \quad (1)$$

where the operator T aligns image $\mathcal{I}_j$ to $\mathcal{I}_i$ based on motion vectors. However, traditional block based motion estimation is difficult to handle the complexity variations among set of similar images, e.g., variations in scale, rotation and illuminance.

In this paper, we take SIFT descriptors to measure the extent and correlation of similar content, which are both used to measure inter-image redundancy. For each images, $\mathcal{I}_j$, the SIFT points should be detected and described with local histogram of gradients. Since the coding efficiency does not have significant difference for smooth area with intra coding or inter coding, we only utilize the SIFT points in the area with large local variations to measure the inter-image redundancy. Then, we refine the SIFT points by deleting those ones, which are in smooth area. To be specific, we discard the SIFT points when the local variance of its neighboring pixels is smaller than a given threshold, e.g., 30 for image with pixel range [0, 255]. In addition, the left SIFT points are further refined by a geometric transform. A homography matrix $\mathbf{H}_{i,j}$, which is derived through the RANSAC [15] method, is used to model the deformation between images $\mathcal{I}_i$ and $\mathcal{I}_j$ and constrain the mapping homogeneity of matched SIFT points. For each pair of matched SIFT points, the deviation of their locations under the model $\mathbf{H}_{i,j}$ is defined as,

$$D(k_{i,j}) = \|\mathbf{x}_i(k_{i,j}) - \mathbf{x}_j(k_{i,j}) \times \mathbf{H}_{i,j}\|^2 \quad (2)$$

where $\mathbf{x}_i(k_{i,j})$ is the coordinate vector, $\mathbf{x}_i(k_{i,j}) = \{x_1, x_2\}$. The matched SIFT points with deviation D larger than a given threshold are also removed. Based on these refinements, most of the SIFT points left are well related with similar content between two images, e.g., in Fig.2.

In order to measure the COS of inter-image redundancy, we define the distance of matched SIFT descriptors as follows,

$$d(f_i(k_{i,j}), f_j(k_{i,j})) = \frac{1}{|\mathbf{u}_i(k_{i,j})|}\|\mathbf{u}_i(k_{i,j}) - \mathbf{u}_j(k_{i,j})\|^2 \quad (3)$$

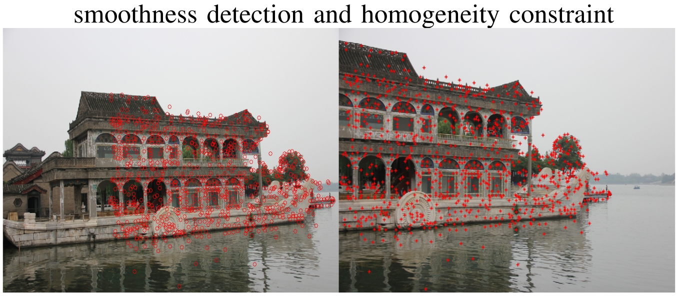
$$\mathbf{u}_n(i) = \frac{\mathbf{v}_n(i)}{\|\mathbf{v}_n(i)\|} \quad (4)$$

where $|\mathbf{u}_i(k_{i,j})|$ is the number of elements in $\mathbf{u}_i(k_{i,j})$. Then, the average distance of all the SIFT points, $\hat{d}_{i,j}$, is utilized to measure the correlation of the similar content between $\mathcal{I}_i$ and $\mathcal{I}_j$.

$$\overline{d}_{i,j} = \frac{1}{N_{i,j}} \sum_{k(i,j)=1}^{N_{i,j}} d(f_i(k_{i,j}), f_j(k_{i,j})) \quad (5)$$

where $N_{i,j}$ is the number of the matched SIFT points between $\mathcal{I}_i$ and $\mathcal{I}_j$.

(a) Matched SIFT points before
smoothness detection and homogeneity constraint



(b) Matched SIFT points after
smoothness detection and homogeneity constraint

Fig. 2. The effect of SIFT point refinement.

In order to measure the EOS of inter-image redundancy, we further introduce the covered area by SIFT points to estimate it. For image $\mathcal{I}_i$ with size of $H_i \times W_i$, its extent of similar content for image $\mathcal{I}_j$ with size of $H_j \times W_j$ is measured by proportion of area covered by matched SIFT points to the whole image area as follows,

$$p_{i,j} = \frac{s_{i,j}}{H_j \times W_j} \tag{6}$$

Here $s_{i,j}$ is the covered area by matched SIFT points in image $\mathcal{I}_j$, which is calculated by adding up the areas of all the triangles generated from the Delaunay triangulation of the discrete SIFT point set. An example is illustrated in Fig. 3. There are three different images (the left two images are the same one actually) in Fig. 3, and the similar content takes up a large proportion to the right image while taking up a small proportion to the left image. Therefore, the left image has much more inter-image redundancy for the other two images than that of the right images.

Based on the two parts, the inter-image redundancy measure for image $\mathcal{I}_i$ is formulated as follows,

$$Q_i = \frac{1}{|S|} \sum_{\substack{j \in S \\ j \neq i}} \left(p_{i,j} + (1 - \overline{d}_{i,j})\right) \tag{7}$$

where $|S|$ is the number of similar images in set S. The larger value of Q_i corresponds to more inter-image redundancy relative to set of other images.

B. Image set compression scheme based on inter-image redundancy measure

Considering the low latency of image access, we take the star topological prediction structure to compress image set, which only has one image delay for the predicted image access. The framework of our image set compression method is illustrated in Fig. 4. The SIFT points are first extracted from every image and then refined by smoothness detection and homogeneity constraint. Second, the distance of matched SIFT descriptors and the covered area are calculated for any two images, then the representative image is selected with the maximum inter-image redundancy. Third, the representative image is compressed with HEVC intra coding method. Finally, the reconstructed representative image is further processed with geometric transform and photometric transform just as in [11] to make it approach the corresponding predicted image. Then, the predicted images are compressed with HEVC inter coding method by referring to their corresponding transformed representative images.

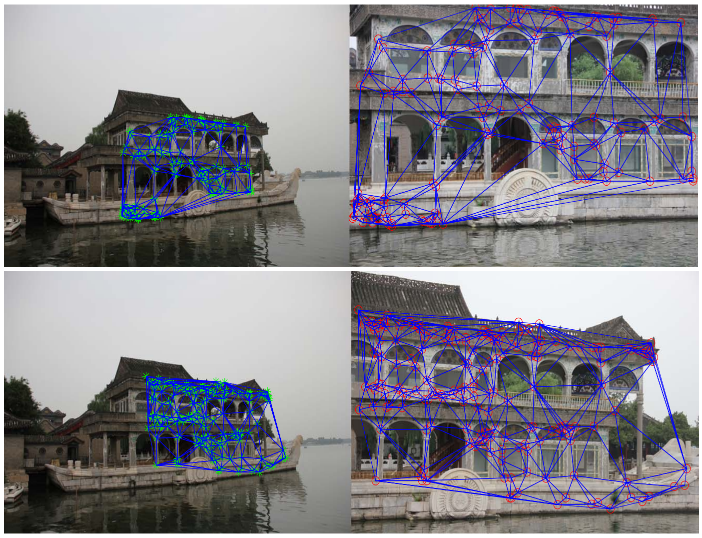


Fig. 3. The prediction area covered by matched SIFT points.

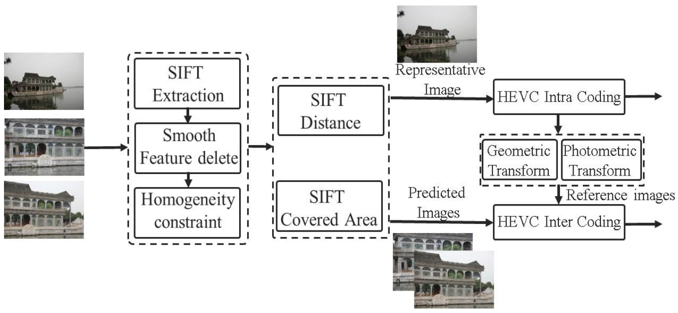
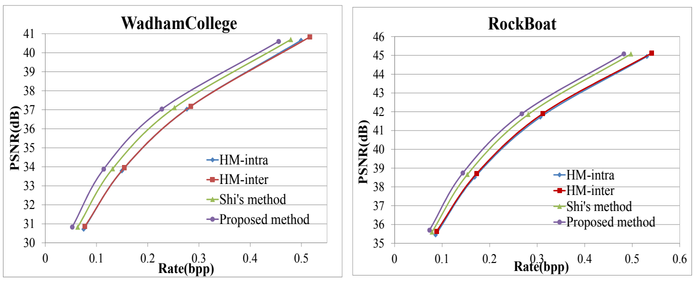


Fig. 4. Image set compression framework in our method.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our proposed inter-image redundancy measure method for image set compression by comparing with state-of-the-art image set compression method (denoted as Shi′s method) in [11], and HEVC intra/inter coding methods. We take the two image sets used in [11], *RockBoat* with 20 images, *WadhamCollege* with 5 images. For each set of images, we first apply our proposed inter-image redundancy measure method and the SIFT distance based method used in Shi′s method to search the representative image. Then, the representative image is compressed with HEVC intra coding method and the other images compressed with HEVC inter coding method by referring to the reconstructed representative image. For our proposed method and Shi′s method, the reconstructed representative image is processed by geometric transform and photometric transform according to the matched SIFT points to generate more approached reference images for the corresponding predicted image. For HEVC inter coding, the reconstructed representative image is directly used as reference image for predicted image coding. For HEVC intra coding, each of the images are compressed individually with HEVC intra coding method.

There are different representative images based on our proposed method and Shi′s method for the two image sets. Fig. 5 shows the rate distortion curves for the two image sets compression with different methods. Compared with other methods, the compression performance with the representative image determined by our proposed method is improved significantly. Table I illustrates the bitrate saving compared with HEVC intra coding method. For image set *WadhamCollege*, our method achieves bitrate saving up to 20.8%. These results show that the representative images determined by our proposed method have more redundancy content for image set and provide better prediction for other images than that in Shi′s method. Therefore, the EOS and COS of similar content together play an important role in measuring inter-image redundancy.



(a) (b)

Fig. 5. Rate-distortion curves for image set compression.

TABLE I. THE BITRATE SAVING COMPARED WITH HEVC INTRA CODING METHOD

Sequence Name	HEVC-inter	Shi′s method	Proposed method
WadhamCollege	0.0%	-11.6%	-20.8%
RockBoat	-1.6%	-10.9%	-16.6%

V. Conclusion

In this paper, we have proposed an inter-image redundancy measure for image set compression. The proposed method measures the inter-image redundancy by jointly estimating the extent and the correlation of similar content between images. The area covered by matched SIFT points and the distance of matched SIFT descriptors are used to estimate the EOS and COS, respectively. We apply the proposed inter-image redundancy measure to determine the representative image for image set compression with star topological prediction structure. Experimental results demonstrate that the determined representative images by our method can provide better prediction for other images and achieve compression performance improvement. In addition, our proposed method is a general inter-image redundancy measure, which is not limited to select a representative image for star topological prediction structure but also can be used to determine coding order with sequential prediction structure.

Acknowledgment

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Prime Ministers Office, Singapore, under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office.

References

[1] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[2] X. Zhang, R. Xiong, S. Ma, and W. Gao, “New image coding scheme with hierarchical representation and adaptive interpolation,” Nov. 2011, pp. 1–4.

[3] G. Wallace, “The JPEG still picture compression standard,” *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, Feb. 1992.

[4] K. Karadimitriou and J. M. Tyler, “Min-max compression methods for medical image databases,” *SIGMOD Rec.*, vol. 26, no. 1, pp. 47–52, Mar. 1997.

[5] X. Zhang, R. Xiong, S. Ma, and W. Gao, “Adaptive loop filter with temporal prediction,” in *Picture Coding Symposium (PCS), 2012*, May 2012, pp. 437–440.

[6] ——, “Adaptive loop filter merge in temporal domain,” *JCT-VC F498, JCT-VC Meeting, Torino*, Jul. 2011.

[7] 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, May 1998.

[8] C.-H. Yeung, O. Au, K. Tang, Z. Yu, E. Luo, Y. Wu, and S.-F. Tu, “Compressing similar image sets using low frequency template,” in *2011 IEEE International Conference on Multimedia and Expo (ICME)*, Jul. 2011, pp. 1–6.

[9] R. Zou, O. Au, G. Zhou, S. Li, and L. Sun, “Image set modeling by exploiting temporal-spatial correlations and photo album compression,” in *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, Dec. 2012, pp. 1–4.

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

[11] ——, “Photo album compression for cloud storage using local features,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 4, no. 1, pp. 17–28, Mar. 2014.

[12] R. Zou, O. Au, G. Zhou, W. Dai, W. Hu, and P. Wan, “Personal photo album compression and management,” in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2013, pp. 1428–1431.

[13] Y. Ling, O. Au, R. Zou, J. Pang, H. Yang, and A. Zheng, “Photo album compression by leveraging temporal-spatial correlations and HEVC,” in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, Jun. 2014.

[14] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[15] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.