DENSE CORRESPONDENCE BASED PREDICTION FOR IMAGE SET COMPRESSION

Yabin Zhang, Weisi Lin, and Jianfei Cai

School of Computer Engineering, Nanyang Technological University, Singapore

ABSTRACT

In this paper, we propose a novel dense correspondence based prediction approach to reduce the inter-image redundancy for image set compression. Unlike previous methods, we manage to utilize the dense correspondence to predict and parameterize the inter-image relation and then reconstruct a new reference for the subsequent HEVC inter-prediction and encoding. Comparing to relevant state-of-the-art feature-based methods, our method is able to locally approximate the inter-image relation and thus more robust to complex local variations. Experimental results show that our proposed approach achieves better coding gains when the local variations are dominant.

***Index Terms*—** Dense correspondence based prediction, image set compression, reference reconstruction, HEVC

1. INTRODUCTION

Over the past few years the image set compression for the cloud-based storage begins to attract more and more attentions. Due to the existence of inter-image redundancy, it's feasible to compress the image set together at a higher ratio than encode them using the individual coding approaches, like JPEG [1] or HEVC intra image coding [2].

At the beginning the samples for image set compression are mostly collected from the specific fields, like medical and satellite imagery. The images are well-aligned and the inter-image variations normally exist in the pixel-to-pixel sense. To reduce the inter-image redundancy the widely used methods are the prediction encoding of the image differences by directly subtracting the images with their references. The problems are mostly concerned with the anchor reference image generation and similar images clustering [3][4].

When it comes to more complex wild image sets like personal photo albums, the direct image subtraction results may be even larger than the original ones. Therefore, the subsequent work mostly adopted the video-like encoding way to reduce the inter-image redundancy using video coding techniques. Au *et al* [5] proposed the inter prediction in video coding with Global Motion Compensation and Local Motion Compensation in the given similar image set, and Zou *et al* [6] proposed Minimal Spanning Tree (MST) encoding structure with depth control to achieve better coding gain for the entire set and fast random access as well. They both formulated the images into video-like sequences and compress them using video inter-prediction coding. However, in their image sets the close temporal correlation and viewpoint changes can be searched out similar to real video sequences. In other image sets with larger variations these methods can be limited.

To deal image sets with larger viewpoint differences like multi-view images, Shi *et al* [7] proposed a multi-model prediction method to explore the inter-image correlation in the major planar regions. With the global geometric deformation and photometric transformation, multiple new reference images are generated from the original reference to better predict the current image. In this way the Block-based Motion Estimation and Compensation (BME & BMC) in HEVC will be more likely to explore and utilize appropriate prediction from the multiple reference images. However, the SIFT-based global photometric transformation is not always effective because the feature keypoints are normally sparse and not sufficient enough for reliable luminance adjustment, and also the global geometric deformation is not always effective to deal with local variations.

In this paper, we propose a novel dense correspondence based prediction approach to achieve robust luminance adjustment and comparable geometric transformation performance comparing to the state-of-the-art. We first adopt the dense correspondence based prediction approach to estimate the consistent pixel-to-pixel relation via the fast-convergent random searching, and then parameterize the correspondence results in the selected units individually. After the parameterization a new reconstructed reference is reconstructed by geometric transformations and luminance adjustments. Finally the inter-prediction and encoding process in HEVC inter coding [2] is utilized to further reduce the local variations with predictions from the reconstructed reference. Since the correspondence in our approach is dense, we are able to transform the reference image in the local units and there are much more matched pixels for the reliable luminance adjustment. Therefore, luminance adjustment results in local regions are more robust than the global transformation.

The rest of the paper is organized as the following: Section 2 gives the overview of our scheme. Section 3 introduces the dense correspondence based prediction and Section 4 presents the new reference reconstruction. Experimental results are shown in Section 5 with performance evaluation and Section 6 is the conclusion and future work.

2. OVERVIEW OF OUR SCHEME

The image set compression can be classified into two parts: image set encoding structure and image pair compression. The encoding structure for the entire image set strongly relies on the image pair encoding approach. As we can find in [6][8], they have chosen different encoding structure generation method because only the similarity metric aligned with image pair compression approach can achieve the optimal encoding structure for the entire set. In this paper we mainly focus on the image pair compression approach to exploit the inter-image redundancy and the encoding structure will be presented in our future work.

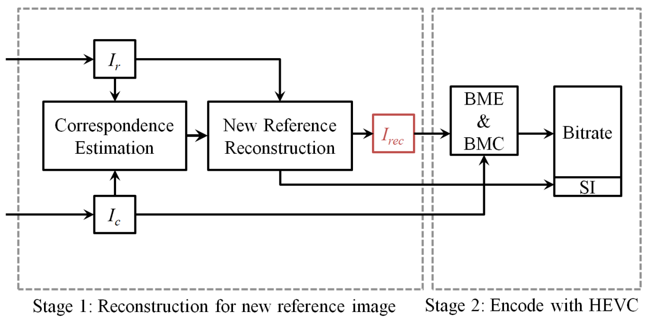

Fig. 1. Framework of image pair compression with dense correspondence based prediction

The framework of our image pair compression approach is shown in Fig 1. The purpose is to encode the current image I_c using less bitrate with the prediction from the reference image I_r in the given image pair. The dense correspondence between I_c and I_r is estimated firstly. Normally it's pretty hard to utilize the dense matching approach like the SIFT flow [9]. These kinds of dense correspondence are non-parameterized and uneconomical as the Side Information (SI) for encoding. In our method, the dense correspondence based prediction between I_c and I_r is constrained to be consistent and thus appropriate for effective parametrization. After the correspondence estimation we estimate the correlation parameters in the non-overlapped units via the widely used RANSAC [10] algorithm. Then the new reference I_{rec} is reconstructed by geometric transformations and luminance adjustments in the selected units for the following BME & BMC in the next stage.

In the second stage, the reconstructed reference I_{rec} is transmitted to the decoded picture buffer for BME & BMC in HEVC. The reconstructed reference works as a better candidate for inter-image prediction. Thus the current image I_c is encoded with the prediction from the reconstructed reference in the same way as video inter-prediction coding and formulated into the bitrate with the necessary parameters as SI. For the I_c decoding, the new reference I_{rec} will be reconstructed again from the reference image I_r and works as the prediction reference for I_c.

3. DENSE CORRESPONDENCE BASED PREDICTION

One of the most important advantages using dense correspondence is that the estimated pixel-to-pixel relation makes it feasible to reconstruct new reference in the local regions, while the sparse feature-based methods are normally limited to the global transformations. Generally, it's impractical to directly apply the BME in video coding to the dense correspondence searching of translations, rotations, and scales in the whole range of reference image due to the huge computation increase. However, if we choose the overlapped blocks division scheme rather than the non-overlapped scheme, the adjacent blocks will gain much stronger correlation and their motion estimation results thus will be similar. The neighborhood coherence plays an important role in the fast-convergent solution for the following searching problem. Here we use the left-top pixels of the blocks to denote the corresponding blocks. Then the overlapped block mapping estimation problem in the range of whole reference image is formulated into the following objective function minimization:

$$\min E = \sum_{mn} MSE(I_c(C_{ij}), I_r(R_{ij})) \tag{1}$$

$$R_{ij} = A \cdot C_{ij}$$

$$A = \begin{bmatrix} scale \cdot cos(\theta) & sin(\theta) & X \\ -sin(\theta) & scale \cdot cos(\theta) & Y \\ 0 & 0 & 1 \end{bmatrix}$$

where E is the total error energy, *MSE* is the Mean Squared Error of the two corresponding blocks, C_{ij} is the block coordinates in the current image I_c at (i, j), R_{ij} is the corresponding block coordinates in the reference image I_r, m and n is the number of rows and columns of all overlapped blocks in I_c, A is the affine matrix including four parameters $(X, Y, \theta, scale)$ describing the mapping relation for each block.

While the problem is still hard to solve, the Generalized PatchMatch (GPM) Algorithm [11] offers the practical, fast-convergent solution with iterations of neighborhood propagation and random searching, but the neighborhood propagated searching results are always distributed in the whole range of I_r and also not consistent due to its greedy collaboratively gradient-descent searching. Here we adopt Yoav Hacohens [12] solution to compromise greedy minimization of E to achieve consistent matching by combining the GPM method with local consistency checks as the following Algorithm 1.

While in the GPM algorithm the natural coherence among overlapped blocks works as weak constraints for the neighborhood propagation and has no guarantee for consistent results, here the additional consistency checks are targeted to remove inconsistent mapping results in neighbor pixels and local regions, and then the left consistent results are kept as the initialized relation map for another round of GPM. By narrowing down the parameter searching ranges at the same

Algorithm 1 GPM with consistency checks

```
GPM initialization
for Parameter range from coarse to fine do
    GPM convergent searching results
    Consistency checks:
        Neighbor pixel consistency check
        Local random region consistency check
        Small regions elimination
    Narrow parameter range (X, Y, θ, scale)
```

time, the following round of GPM always produces more consistent results in the region sharing similar content. At last we keep the final consistent results as the prediction results for the new reference reconstruction as shown in Fig 2 (a) and (b). For the implementation details of the dense correspondence with consistency checks, the reader can refer to [12].

4. NEW REFERENCE RECONSTRUCTION

We utilize the square unit division scheme to parameterize the consistent prediction results adaptively in the local region. By partitioning the matched pixels of the current image I_c into regular non-overlapped units, the correlation in each unit pair is approximated by using homograph transformation, and the subsequent luminance adjustment is used to balance the luminance variations in each unit. The new reference image is thus reconstructed with the selected units transformed from the original reference image. The reference reconstruction process is shown in Fig 2.

The new reconstructed reference I_{rec} has the same resolution as I_c and the coordinates are partitioned into regular non-overlapped 256×256 units from the left-top. To balance the redundancy reduction and SI, we choose the multiple of the maximal coding unit size in HEVC as the basic unit size and in our experiments the multiple is 4. The units with a small percentage of matched points are normally not reliable and worthy to reconstruct for prediction and here the unit selection threshold we adopted is $\tau = 0.2$ empirically. Then we exploit the general RANSAC algorithm [10] to generate the homograph transform H for the approximation of the geometric relation in each of selected units in Equation 2.

$$u_r = u_c \cdot H, \qquad H = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \tag{2}$$

where u_c and u_r are the matched point coordinates in each unit from I_c and I_r. The homograph transform H is used to approximate the inter-unit relation between the image pair by capturing the major deformation relation as shown in Fig 2 (c) and (d). After the geometric correlation parameterization, the geometric transformation and luminance adjustment in each unit are used to reconstruct new reference I_{rec} as shown in the following Equation 3:

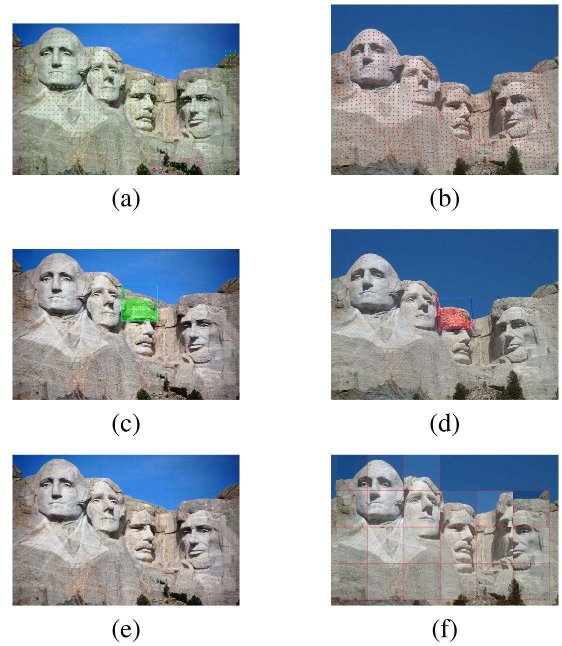

Fig. 2. Dense correspondence based prediction and new reference reconstruction. (a) and (b) Matched pixels in I_c and corresponding points in I_r. (Note that the matched pixels in (a) and (b) are partially shown every 32 pixels.) (c) and (d) One of the selected units in I_c and its corresponding unit in I_r. (e) Current image I_c. (f) Reconstructed reference I_{rec}.

$$\begin{aligned} I_{rec}(U_c) &= \alpha \cdot I_r(U_c \cdot H) + \beta \\ \alpha &= \sigma(I_c(U_c))/\sigma(I_r(U_c \cdot H)) \\ \beta &= \mu(I_c(U_c)) - \alpha \cdot \mu(I_r(U_c \cdot H)) \end{aligned} \tag{3}$$

where I_c and I_r are the current and reference images, and I_{rec} is the reconstructed reference image. U_c are coordinates of all pixels in selected unit in I_c and $U_c \cdot H$ are their corresponding coordinates in I_r, μ and σ are the mean and variance of the pixel values, α and β are the luminance scale and offset. Here we only perform the luminance adjustment since the luminance component is more important in HEVC with 4:2:0 color space sampling. Next the new reference I_{rec} is reconstructed in the selected transformed units and the left uncovered regions are filled with I_r as shown in Fig 2 (f).

At last we utilize the HEVC Inter-prediction coding to encode I_c with the reference I_{rec}. Here the SI consists of selected unit locations, homograph matrices, and luminance adjustment parameters. The matrices are quantized and rounded with the appropriate constant quantization table. The locations of selected units and luminance parameters are also quantized and round properly. Finally the SI is encoded with arithmetic coding and kept along with video encoded bitrate.

Fig. 3. Test image pairs
(Top row: image pairs from "Mail room" and "Wadham College";
Bottom row: image pairs from "Notre Dame" and "Mount Rushmore")

5. EXPERIMENTAL RESULTS

For the comparison, we re-implement the SIFT-based multi-model method in [13]. We utilize the HEVC HM15.0 for both methods and choose the Quantization Parameter (QP) setting of 22, 27, 32 and 37. The experiments are conducted in four image sets: "Mail room" set, "Notre Dame" set, "Wadham college" set provided in [13] and "Mount Rushmore" set collected from Google Images as shown in Fig 3.

The coding results are shown in Fig 4. The SIFT-based multi-model method saves 31.57%, 51.76%, 38.63%, and 9.05% over the HEVC inter-prediction coding in the four image pairs while our approach achieves 43.63%, 55.60%, 34.20%, and 18.73% bitrate saving. We can observe that our approach achieves better coding performance for the image pairs in Fig 4 (a) and (b). In Fig 4 (c) the performance is comparable as the state-of-the-art due to the existence of obvious multi-view related planar regions and side effect of the unit division for single reference reconstruction. It's obvious that the inter-image coding improvement is content-dependent. Although in Fig 4 (d) the images collected from Google Images are not as similar as other three sets, the improvement of our approach is comparatively significant and demonstrates that our local luminance adjustments are more robust comparing to the feature-based global photometric transformation when the local variations are dominant and luminance relation is complex.

Comparing to the state-of-the-art methods, our approach achieves better performance and robustness to complex luminance variations when the local variations are dominant. Generally, the feature-based image pair method has no strong relation to the inter-image coding prediction. While our dense correspondence based prediction approach is similar to BME & BMC in video coding, it has the potential to more exactly estimate the encoding improvement, and for the entire image set the better prediction method aligned with image pair compression approach means the better coding structure and thus better image set encoding performance.

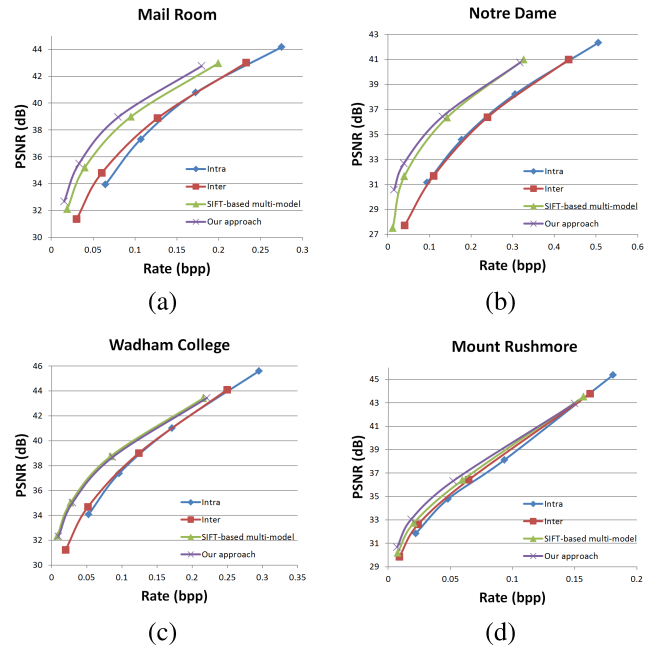

Fig. 4. Coding performance comparison. ("Intra" and "Inter" curves are the HEVC Intra-prediction coding and Inter-prediction coding, and "SIFT-based multi-model" [13].)

6. CONCLUSION

As a response to the new requirement and challenge in big data and cloud computing, we have proposed a novel dense correspondence based prediction approach for image set compression which achieves better performance comparing to the relevant state-of-the-art method when the local variations are dominant. The dense correspondence is first estimated to predict the pixel-to-pixel relation via fast-convergent random searching, and with the parameterization of the prediction results, we reconstruct a new reference image by geometric transformations and luminance adjustments for the subsequent HEVC inter-prediction coding. Because of the reconstruction in local units our approach is more robust to the complex local variations. In the image set compression that follows in our future work, the dense correspondence based prediction has the potential to better estimate the image pair similarity for the optimal encoding structure generation and thus achieves better performance for the entire image set.

7. 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

8. REFERENCES

[1] Gregory K Wallace, "The JPEG still picture compression standard," *Consumer Electronics, IEEE Transactions on*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[2] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay 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.

[3] Kosmas Karadimitriou and John M Tyler, "The centroid method for compressing sets of similar images," *Pattern Recognition Letters*, vol. 19, no. 7, pp. 585–593, 1998.

[4] Anthony Schmieder, Howard Cheng, and Xiaobo Li, "A study of clustering algorithms and validity for lossy image set compression.," in *IPCV*, 2009, pp. 501–506.

[5] Oscar Au, Sijin Li, Ruobing Zou, Wei Dai, and Lin 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*. IEEE, 2012, pp. 84–90.

[6] Ruobing Zou, Oscar C Au, Guyue Zhou, Wei Dai, Wei Hu, and Pengfei Wan, "Personal photo album compression and management," in *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*. IEEE, 2013, pp. 1428–1431.

[7] Zhongbo Shi, Xiaoyan Sun, and Feng Wu, "Multi-model prediction for image set compression," in *Visual Communications and Image Processing (VCIP), 2013*. IEEE, 2013, pp. 1–6.

[8] Zhongbo Shi, Xiaoyan Sun, and Feng Wu, "Feature-based image set compression," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[9] Ce Liu, Jenny Yuen, and Antonio Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 5, pp. 978–994, 2011.

[10] Martin A Fischler and Robert 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, 1981.

[11] Connelly Barnes, Eli Shechtman, Dan B Goldman, and Adam Finkelstein, "The generalized patchmatch correspondence algorithm," in *Computer Vision–ECCV 2010*, pp. 29–43. Springer, 2010.

[12] Yoav HaCohen, Eli Shechtman, Dan B Goldman, and Dani Lischinski, "Non-rigid dense correspondence with applications for image enhancement," in *ACM Transactions on Graphics (TOG)*. ACM, 2011, vol. 30, p. 70.

[13] Zhongbo Shi, Xiaoyan Sun, and Feng Wu, "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, March 2014.