Compression with Vision Technologies

Xiaoyan Sun, Feng Wu, Shipeng Li

Microsoft Research Asia, Beijing, China, 100081

Abstract. In this paper, we propose a method for image compression in which currently developed vision technologies are incorporated with traditional transform-based coding methods. Based on newly developed vision technologies, we present two coding-oriented inpainting methods: structure-aware inpainting and object-based inpainting. Examples show the effectiveness of these inpainting methods on information recovery employing certain kinds of assistant information. Moreover, a primary image compression system has been built to test the proposed compression approach in which some regions are intentionally removed at the encoder only to be restored naturally at the decoder. Up to a 50% bit savings can be achieved at similar visual quality levels when compared with JPEG.

Index Terms—Image compression, image coding, image inpainting, visual quality, perceptual quality

1. INTRODUCTION

Over the past two decades, attempts have been made to develop compression techniques by identifying and utilizing features within images to achieve high coding efficiency [1]. Moreover, the human visual system (HVS) [2][3] is incorporated into coding methods, trying to remove some of visual redundancy inherent in images and better the visual quality of resulting images. Nevertheless, the development of such coding schemes is greatly influenced by the availability as well as effectiveness of some algorithms, such as edge detection and segmentation tools.

Meanwhile, great improvements have been made in traditional signal-processing-based compression methods. Such mainstream coding schemes manage to make use of the statistical redundancy among pixels in pursuit of high coding efficiency. State-of-the-art JPEG2000 and MPEG-4 AVC/H.264 are two such examples that greatly outperform previous generations in coding efficiency. However, perceptual quality is almost neglected during algorithm design. In addition, current development of such schemes shows that even small improvements are commonly accomplished at the pain of multiplying encoding complexity.

Recently, vision-related technologies have shown remarkable progress in hallucinating pictures with good perceptual quality. Essentially, image compression schemes and vision systems face a similar problem, that is, how to represent visual objects in efficient and effective ways. Is it possible to apply certain vision technologies in compression systems aiming toward perceptual quality rather than pixel-wise fidelity?

In fact, attractive results have been achieved by newly presented vision technologies, such as object segmentation, feature extraction, image completion and texture synthesis [4]~[6]. It is promising to significantly further reduce visual redundancy on the basis of current transform-based coding schemes, which has been exemplified by the success of applying image inpainting technologies in error concealment [7][8]. However, compression systems would greatly benefit if vision methods are introduced with regard to compression. In other words, as the complete source images are available in compression systems, new vision technologies could be proposed for all available source signals to be fully exploited.

In this paper, a pilot study on compression using vision technologies is presented. First, we introduce several vision algorithms that could prove beneficial to efficient image compression. Second, the image compression system brings new challenges as well as opportunities for developing new computer vision approaches since the coding system provides different requirements and conditions. Indeed, two compression-oriented vision methods - structure-aware inpainting and object-based completion - are presented to make use of all available source information. Then, we outline the image compression system based on presented vision approaches. We provide a testing system and experimental results to show the effectiveness of the vision-assisted compression scheme. Compared with JPEG, our proposed image compression scheme has the potential of up to a 50% bit-saving capability at similar visual quality levels.

The rest of this paper is organized as follows. Several related computer vision technologies and our proposed compression-oriented inpainting approaches are described in Section II. Then, Section III presents the image compression system based on the proposed vision technologies. Section VI concludes the paper and discusses proposed coding system issues.

2. ASSISTED VISION TECHNOLOGIES

Recently, state-of-the-art vision-related technologies have shown remarkable progress in hallucinating pictures that have adequate perceptual quality. Some of these vision technologies can be readily employed by image compression systems to improve coding performance in terms of perceptual quality.

2.1 Available vision technologies

With the rapid development in computer vision and multimedia applications, there are many advanced vision approaches that would clearly bring profit to image and video compression. Examples include object segmentation, image completion, edge detection, and so on.

2.1.1 Object-based Approaches

Object-based schemes have been widely exploited in image and video coding. Improvement on the segmentation quality would clearly accelerate the development of the object-based schemes. In the past several years, a lot of semi-automatic methods [9][10] have been proposed to decipher objects by user defined features, contours or regions. Fully automatic methods [11] based on specific characteristics have also been presented. Results are quite encouraging for object-based coding approaches, as shown in the literature [12][13].

2.1.2 Region-growing Methods

It is accepted that redundancy inside the image can be removed by intra block prediction in current image coding schemes. Such approaches are efficient for flat area coding. However, for textural regions, such as lawns, brick walls, wallpaper, and so on, it is hard for intra-prediction methods to achieve coding efficiency or visual quality.

While pixel-wise fidelity is not vital, perceptual redundancy as well as the statistical redundancy is desirable for extraction in image compression. Texture synthesis, which can generate a newer form of output from the presence of texture, is a good candidate for image compression to deal with such textural regions. There are many examples showing that, when presented with a specified small textual sample image, texture synthesis is able to produce a large image with perceptually similar texture [14][15].

Moreover, image inpainting (also known as image completion) is a process of restoring missing data in a designated region of an image in a visually plausible way. Current results of image inpainting illustrate that it can naturally recover homogenous regions as well as certain kind of structural regions [6][16].

Therefore, image inpainting, as well as texture synthesis, can be incorporated into current image coding systems to eliminate perceptual redundancy. In such system, only image portions or samples must be encoded and transmitted. At the decoder side, the whole image can be restored through image inpainting or texture synthesis. The method is promising in achieving high compression ratio at similar visual levels.

2.1.3 Structure-related Techniques

In image processing, structural information, especially edge, has a perceptual significance that is greater than the numerical value of the energy contribution to the entire image. Coding efficiency, as well as video quality, of image coding approaches can be improved based on desired structural information.

Traditional edge detectors are based on local properties of image variation. Today, gradient based edge detection is the one most frequently used [17]. Recently, proposed generation [18] has integrated image segmentation into edge detection, making a good trade-off between the sensitivity of homogenous region delineation and the over segmentation of the image.

Additionally, sketch is another efficient descriptor of an image [19]. It refers not only to the geometry of the edges but also to the relevant gray-level information across them. This makes sketch a rich representation of an image. Examples reported currently show that encouraging results can be achieved for a large number of generic images [20][21]. Thus, such sketch modeling algorithms would directly lead to image coding schemes.

There are many other vision techniques that can be employed in coding schemes to provide better coding performances. Examples include, but are not limited to, algorithms for image hallucinating [22], shape matching [23], analogy [24], epitome [25] and textons [26].

2.2 Compression-oriented Methods

On the one hand, we can straightforwardly use existing vision technologies in compression; on the other, it should be noticed that new task-driven vision technologies could be developed for the sake of compression. So far, we have done some research on compression-oriented image inpainting method.

There is a common assumption in image inpainting that structure and texture information in unknown regions can be directly or indirectly derived from surrounding known regions. But for more general cases, unknown regions may contain structure information that is hardly inferred even by human intelligence. In typical inpainting applications (e.g. deterioration and element recovery), no existing approaches can tackle this problem because of the lack of relevant information.

The significant difference in the compression scenario is that the whole image is available. Hence, many kind of assistant information can be extracted from the source image to help allow the inpainting. Here, two approaches - structure-aware inpainting and object-based image inpainting - have been proposed, utilizing different kinds of assistant information.

2.2.1 Structure-aware Inpainting

We propose a structure aware inpainting (SAI) technique in [27], where two main problems are addressed.

The first is: What should be extracted as the structure information from the image? From image representation in computer vision, we derive that the structure information can be described by one-pixel-width binary curves. It can be efficiently compressed with arithmetic coding.

The second is: How do we complete the unknown regions with available binary structure information? It is not a simple problem because structure information may be complex. A distance-based inpainting algorithm is proposed to propagate the structure. Then, the rest part of unknown region is restored by texture synthesis.

Fig. 1: Image inpainting with complex structure. (a) Original image; (b) Unknown region (denoted by black blocks); (c) Extracted structure (denoted by white lines); (d) Completed image

According to structure-related properties, including colors of the area delimited by the structure, distances of pixels with respect to the structure, and variance of the conjoint structures, pair matching integrated with pixel interpolation is proposed for structure propagation.

Fig. 1 is an example of our proposed structure propagation scheme. Most portions of the words are removed from the original image, while the proposed scheme can still naturally reconstruct them.

2.2.2 Object-based Inpainting

We also propose an object-based approach to address inpainting problems in much complicated scenarios, a continuing challenge in even state-of-the-art techniques. In such cases, missing regions may contain complex structures as well as parts of objects or entire objects. This makes it difficult to propagate missing regions from neighboring regions.

When the target and the source objects are specified, with the objects selected either manually or automatically, we perform an object inpainting algorithm in three steps. First, we roughly match source object and target object based on some key features by two-dimensional (2D) geometrical operations, such as translation, rotation, and scaling. Then, the structure of the object is warped from source to target by pixel-level deformation. A patch-based texture synthesis method is used to restore the holes between the attached texture and structure. Moreover, an automatic merging method is presented to naturally integrate the rendered object into the target image.

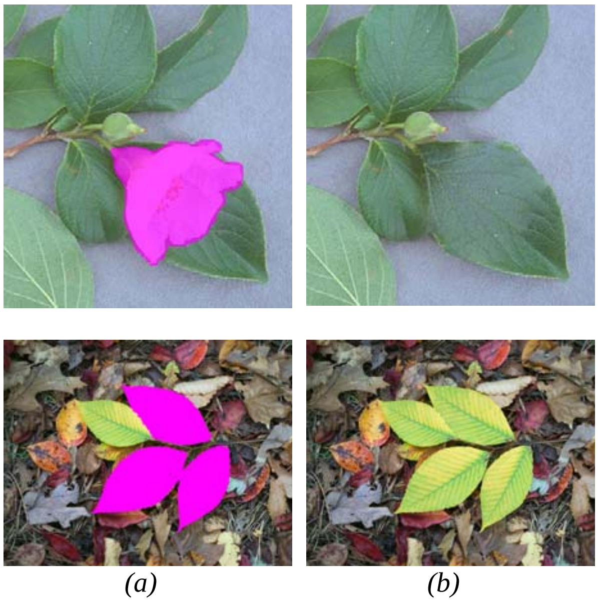


Fig. 2: Results of object-based image inpainting. (a) Original image with unknown region (magenta); (b) recovered image by object-based inpainting.

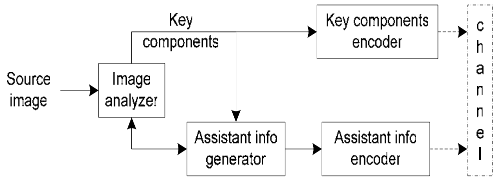
Examples, as shown in Fig. 2, demonstrate the effectiveness of our method in object completion. It is able to recover a species of complex objects in a visually pleasant manner.

3. PROPOSED COMPRESSION SYSTEM

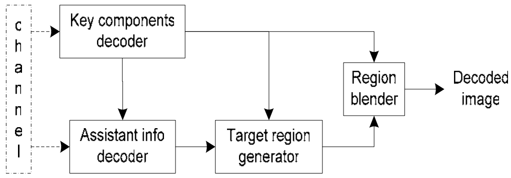
Realizing the capabilities of vision technologies in compression, we propose an image coding approach by taking advantage of compression-oriented vision tools. A prime system is built based on our proposed inpainting methods.

3.1 Framework of the Coding System

The framework of our proposed image compression system is depicted as in Fig. 3.



(a) Encoder diagram



(b) Decoder diagram

Fig. 3: Block diagram of proposed compression system.

At the encoder side, the image analyzer, together with the assistant information generator, separates the source image into key components and droppable regions. Meanwhile, the assistant info generator extracts the image edges as the structure info in the droppable regions. In the following inpainting and coding processes, key components are considered as known regions while droppable regions are regarded as unknown regions.

Given the key components and the assistant edge information, key components are coded as a common image by the traditional transform-based image coding method, such as JPEG; the assistant edge information is organized as binary structure curve and coded by binary arithmetic coding.

As shown in Fig. 3 (b), the key components as well as the assistant edge information are first decoded. Then, using our proposed inpainting method, regions skipped at the encoder will be restored. Finally, recovered regions are integrated with the key components to generate the entire image by the region blender.

3.2 Experimental Results

Experimental results show the potential of our proposed vision-based approach in image compression. The edge detection method in [18] is used in our current system to extract the structure information.

Fig. 4 shows some results generated by our system. In this figure, the first column indicates the decoded image coded by JPEG. The second column shows the resulting image generated by our proposed coding system.

As illustrated in Fig. 4, the visual quality of the decoded image generated by our scheme is quite similar to that of a JPEG coded image, and has a significant bits savings. Up to 55% savings is achieved.

JPEG compressed (64.35kB)

Proposed scheme (28.49kB)

JPEG compressed(147.44kB)

Proposed scheme (104.62kB)

Fig. 4 Visual quality evaluation on the reconstruction images of our proposed compression scheme in compared with JPEG.

4. DISCUSSION AND FUTURE WORK

In this paper, a vision-based approach for image compression is proposed by integrating the state-of-the-art vision technologies into current coding methods. Several vision algorithms are enumerated as promising candidates. Moreover, we present two coding-oriented inpainting methods. Examples prove the effectiveness of the schemes in information recovery using certain kinds of assistant information.

Additionally, a prime image compression system we built tests the proposed method. In this system, some regions are intentionally removed at the encoder and can be naturally restored at the decoder. The potential of the proposed coding approach has been evaluated in reference to JPEG coding scheme. Significant bit-savings - up to 50% - can be achieved at similar visual quality levels when compared with JPEG.

The proposed method is far from mature, however. There are many aspects that need to be investigated at greater depth and improved.

First, how to classify the regions is still a challenge in the proposed compression scheme. A semi-automatic segmentation method is used in our current system, although it is not suitable for compression. Second, the proposed inpainting method is based on some key features that may be hard to effectively extract in the case of complex images. Moreover, the robustness of the proposed scheme should be improved. We would like address these questions in the future.

ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance of Dong Liu and Chen Wang in the programming and computational work. Many thanks to Harry Shum and Jian Sun for all their helpful discussions and to Dwight Daniels for his editing.

REFERENCES

1. M. M. Reid, R. J. Millar, and N. D. Black, "Second-generation image coding: an overview," ACM Computing Surveys, Vol. 29, No. 1, March 1997.
2. I. Höntsch and L. J. Karam, "Locally adaptive perceptual image coding," IEEE Trans. Image Processing, vol. 9, pp. 1472–1483, Sept. 2000.
3. J. A. Saghri, P. S. Cheatham, and A. Habibi, "Image quality measure based on a human visual system model," Opt. Eng., vol. 28, no. 7, pp. 813–818, 1989.
4. M. Bertalmío, G. Sapiro, V. Caselles and C. Ballester, "Image inpainting," SIGGRAPH 2000, Computer Graphics Proceedings, July 2000.
5. V. Kwatra, A. Schödl, I. Essa, G. Turk and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," SIGGRAPH 2003, Computer Graphics Proceedings, July 2003.
6. J. Sun, L. Yuan, J. Jia and H. Shum, "Image completion with structure propagation," SIGGRAPH 2005, Computer Graphics Proceedings, July 2005.
7. S.D. Rane, G. Sapiro and M. Bertalmio, "Structure and texture filling-in of missing image blocks in wireless transmission and compression," Image Processing 2002, International Conference on, 2002.
8. Y. Chen, X. Sun, F. Wu, Z. Liu and S. Li, "Spatio-temporal video error concealment using priority-ranked region-matching", International conference on image processing, vol. 2, pp 1050-1053, 2005
9. A. Cavallaro and T. Ebrahimi, "Object-based video, extraction tools, evaluation metrics and applications," in Proc. of SPIE VCIP, 2003
10. Y. Li, J. Sun, C. K. Tang and H. Y. Shum, "Lazy snapping," in Proc. of SIGGRAPH'04, pp. 303-308, 2004,
11. C. Zitnick, S. Kang, M. Uyttendaele, S. Winder and R. Szeliski, "High-quality video view interpolation using a layered representation," ACM Trans. On Graphics, 2004
12. M.S. Schmalz and G.X. Ritter. "Region segmentation techniques for object-based image compression", Proc. SPIE 5561:62-75, 2004.
13. Q. Wu, K.-T. Ng, S.-C. Chan, and H.-Y. Shum, "An object-based compression system for a class of dynamic image-based representations," Proc. of SPIE Vol. 5960: 1286-1294, 2005.
14. L. Liang, C. Liu, Y.-Q., Xu, B. Guo, and H.-Y, Shum, "Real-time texture synthesis by patch-based sampling," ACM Trans. Graph. 20, 3, 127–150, 2001.
15. S. Lefebvre and H. Hoppe, "Parallel controllable texture synthesis," ACM Trans. Graph. 24, 3, 777–786. 2005.
16. A., Criminisi, P. Pérez, and K, Toyama, "Region filling and object removal by exemplar-based image inpainting". IEEE Trans. on Image Process. 13, 9, 1200–1212. 2004
17. Y. Itoh, An edge-oriented progressive image coding, IEEE Trans. Circuits Systems Video Tech. 6, 2, 135-142, April 1996
18. P. Meer and B. Georgescu, "Edge detection with embedded confidence," Pattern Anal. Machine Intell., IEEE Trans. on, Vol. 23, pp. 1351-1365, Dec. 2001.
19. S. Carlsson, Sketch based coding of grey level images, Signal Processing, 15, 57-83, July 1988.
20. A. Chalechale, G. Naghdy, and A. Mertins, "Sketch-Based Image Matching Using Angular Partitioning," IEEE trans. On systems, man, and cybernetics-part A: systems and humans, Vol. 35, No. 1, Jan. 2005
21. C. Guo, S.-C. Zhu, and Y. N. Wu, "Primal Sketch: integrating texture and structure," http://repositories.cdlib.org/uclastat/papers/2005032101
22. J. Sun, N.-N. Zheng, H. Tao, and H.-Y. Shum, "Image hallucination with primal sketch priors," Proc. of 2003 Conference on Computer Vision and Pattern Recognition (CVPR'03), 729-736, 2003
23. S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.24, No. 24, 509-522,2002
24. A. Hertzmann , C. E. Jacobs , N. Oliver , B. Curless ,and D. H. Salesin, "Image analogies," Proc. of the 28th annual conference on Computer graphics and interactive techniques, p.327-340, August 2001
25. N. Jojic, B. Frey, and A. Kannan, "Epitomic analysis of appearance and shape," in ICCV 2003, vol. 1, pp. 34–41.
26. S. C. Zhu, C. E. Guo, Y. Wu, and Y. Wang. "What are textons," In Proc. Of European Conf. on Computer Vision (ECCV), 2002.
27. C. Wang, X. Sun, F. Wu,†and H. Xiong, "Image compression with structure-aware inpainting," Proc. of ISCAS, 2006