Local Feature Based Salient Region Detection

Chao Lv
Parallel Processing Institute
Fudan University
lch@fudan.edu.cn

ABSTRACT

Local feature descriptors have become the most important part in image / video retrieval systems. But considering the great amount of local features, which could be thousands of local features in one HD photo, it's hard to compute them efficiently in a realistic system. In our research, we try to overcome this obstacle with a straightforward local feature reduction processing by using salient region detection. Furthermore, we also present a efficient method to detect multi salient regions. The whole algorithm is only based on the local features, which means almost no additional overhead involved in our algorithm. In our tests, we compare the LFSR (Local Feature based Salient Region) algorithm with a state-of-the-art algorithm. And the LFSR shows a thousands of times speedup in runtime with acceptable precision and recall loss. When integrated with the SURF descriptor, LFSR can provide a overall 1.6X speedup for the whole processing [**?**].

1. INTRODUCTION

Our society has entered a data-centric era and a huge amount of data are transferred and processed on the Internet. Among them, multimedia data, such as image and video, has become one of the major data types being processed. As analyzed by CISCO Inc., video data occupies 50% of network traffic in 2011 and will increase to 90% in 2013 [1]. According to a report [5], as one of the most popular video sharing sites, more than 20-hour new videos are uploaded to *YouTube* every minute. Moreover, as two most popular photo sharing sites, *Facebook* and *Flickr* host billions of user-uploaded images respectively.

With the rapid increase of multimedia data, one of the most significant challenges is to understand and interpret such a huge amount of multimedia data. Currently, more and more retrieval applications are emerging to process these multimedia data, such as video recommendation [9], travel guidance systems [4] and content-based TV copy identification [6]. In these systems, a fundamental step is to efficiently retrieve useful information from multimedia data. To guarantee retrieval accuracy, typical applications usually extract and utilize hundreds of features to represent an image or a video frame. Thus, unlike traditional text-based retrieval applications, multimedia retrieval applications are not only more data-intensive but also more computation-intensive, which lead to great pressure on real-time processing. For example, SIFT [7] and SURF [3] are two most widely-used image retrieval algorithms [8] [2]. When executed on general-purpose processors, they can only achieve a process speed of about three images per second.

2. CONCLUSION

Acknowledgement

3. REFERENCES

[1] Cisco Visual Networking Index: Forecast and Methodology, 2010-2015. 2011.

[2] J. Bauer, N. Sunderhauf, and P. Protzel. Comparing Several Implementations of Two Recently Published Feature Detectors. *International Conference on Intelligent and Autonomous Systems*, 2007.

[3] H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. *European Conference on Computer Vision*, pages 404–417, 2006.

[4] Y. Gao, J. Tang, R. Hong, Q. Dai, T. S. Chua, and R. Jain. W2GO: a travel guidance system by automatic landmark ranking. *ACM Multimedia*, pages 123–132, 2010.

[5] C. Jansohn, A. Ulges, and T. M. Breuel. Detecting pornographic video content by combining image features with motion information. *International Conference on Multimedia*, 2009.

[6] A. Joly, C. Frelicot, and O. Buisson. Robust Content-Based Video Copy Identification in a Large Reference Database . *International Conference on Image and Video Retrieval*, pages 511–516, 2003.

[7] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[8] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[9] B. Yang, T. Mei, X.-S. Hua, L. Yang, S.-Q. Yang, and M. Li. Online video recommendation based on multimodal fusion and relevance feedback.

International Conference on Image and Video Retrieval, pages 73–80, 2007.