

Next “Generation” Photo and Video Technology

Xiaoou Tang

*Microsoft Research Asia
Beijing, P. R. China*

It is fundamentally important to develop long-term high-impact technology that aims at future applications. It is also important to develop technology that can make immediate impact on our daily life. In general, it is not difficult to build a research demo in computer vision research. However, it is difficult to build a vision system that works in real life. As a result, the impact of computer vision on our daily life is relatively limited compared to other research areas, such as computer graphics, multimedia communication, and networking. In this talk, I will discuss some practical vision topics that we are working on at Microsoft Research Asia. In particular, I would like to demonstrate some technologies that are simple to use, and perhaps more importantly are fun to use. Most of the works focus on processing of children’s photo. We hope these technologies are useful for processing of photos and videos of our children (the next generation of us) now instead of after another generation.

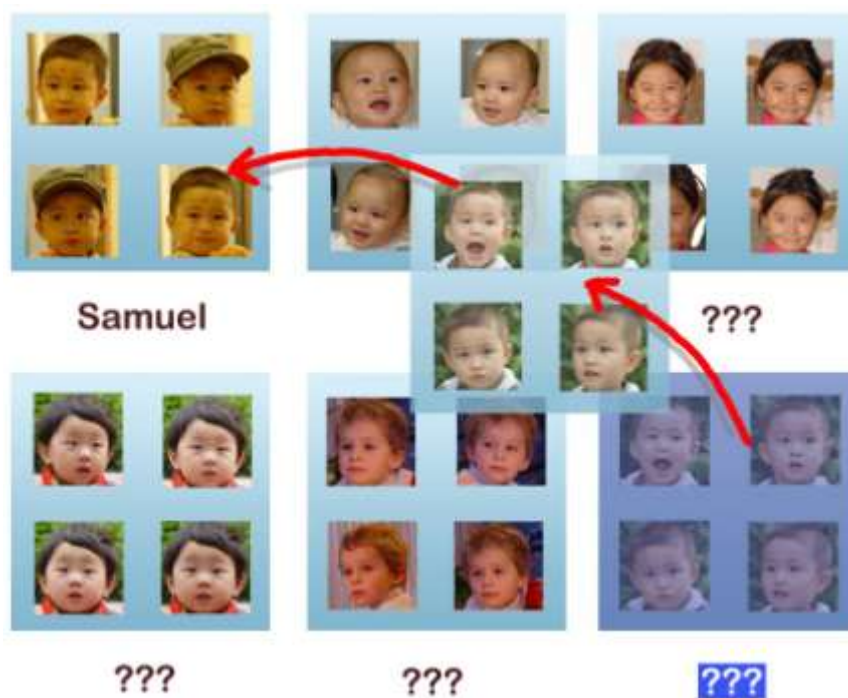


Figure 1. Clustering based photo annotation. Assisted by face detection, recognition, and clustering technology, similar faces are put together into clusters, based on which efficient annotation can be carried out in a file explorer like interface by easy interactions such as drag and drop [1][2].

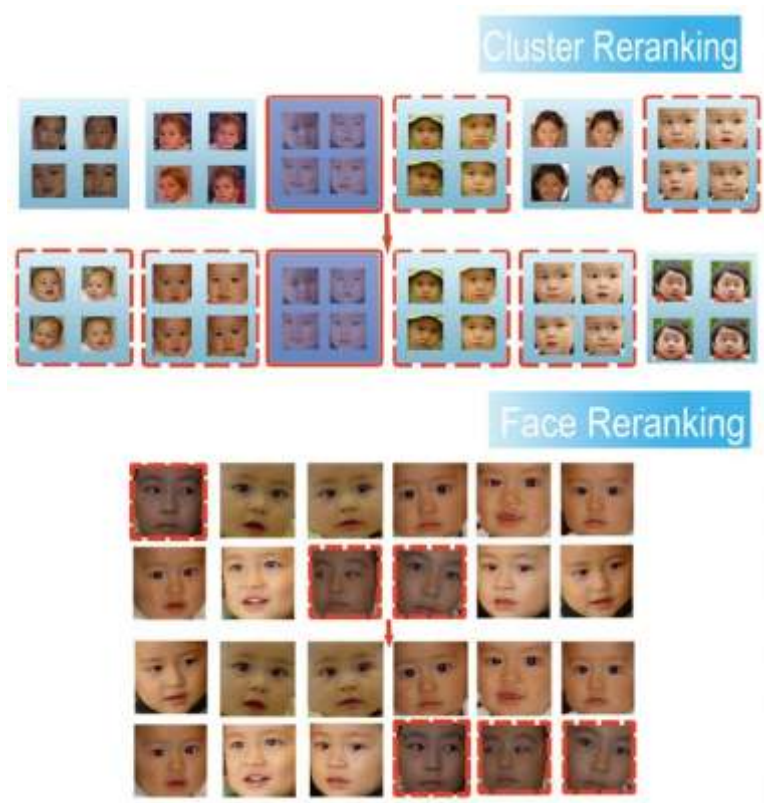


Figure 2. By making use of the tagged faces and analyzing underlying distribution of untagged faces, we can re-rank the spatial configuration of clusters and faces inside each cluster in our user interface to better facilitate annotation [1][3].



Figure 3. Present the photo search results using picture collage [4]. Similarly, video clips can also be presented as a video montage [5].



Figure 6. Using 3D face tracking [8] and background cut technology [9] we add special digital effect to both the foreground face and the background scene of the live video during MSN online video chat.

References:

- [1] Jingyu Cui, Fang Wen, Rong Xiao, Yuandong Tian, and Xiaoou Tang, EasyAlbum: An Interactive Photo Annotation System Based on Face Clustering and Re-ranking, Proceedings of the 2007 Conference on Human Factors in Computing Systems, SIGCHI 2007, ACM Press, 2007
- [2] Xiaogang Wang and Xiaoou Tang, A unified framework for subspace face recognition, IEEE Transaction on Pattern Analysis and Machine Intelligence, 26(9):1222–1228, 2004.
- [3] Yuandong Tian, Wei Liu, Rong Xiao, Fang Wen, and Xiaoou Tang, A Face Annotation Framework with Partial Clustering and Interactive Labeling, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [4] Jingdong Wang, Jian Sun, Long Quan, Xiaoou Tang, and Heung-Yeung Shum, Picture Collage, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006.
- [5] Hongwen Kang, Yasuyuki Matsushita, Xiaoou Tang, and Xuequan Chen, Space-Time Video Montage, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006.
- [6] Jian Sun, Jian Sun, Sing-Bing Kang, Zongben Xu, Xiaoou Tang, and Heung-Yeung Shum, Flash Cut: Foreground Extraction with Flash/No-Flash Image Pairs, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [7] Tie Liu, Jian Sun, Nanning Zheng, Xiaoou Tang, and Harry Shum, Learning to Detect A Salient Object, Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, 2007.
- [8] Qiang Wang, Weiwei Zhang, Xiaoou Tang, and Harry Shum, Real-Time Bayesian 3-D Pose Tracking, IEEE Transactions on Circuits and Systems for Video Technology, 16(12):1533-1541, 2006.
- [9] Jian Sun, Weiwei Zhang, Xiaoou Tang, and Heung-Yeung Shum, Background Cut, Proceedings of European Conference on Computer Vision, 2006.