Denoising face recognition via clustering techniques

Trung Vu, Sean Lee, Alexandre Lamy

Abstract

In this paper we address the problem of detecting and recognizing specific objects (such as the faces of a specific group of people) in real life videos. By using unsupervised clustering techniques, we are able to exploit the inter-frame relationships in the video to significantly improve the accuracy of basic object detection and recognition algorithms. This yields results which are also much more robust to frame switches, lengthy occlusions, and variable numbers of targeted objects leaving and reentering the frames (all common occurrences in real life videos) than standard tracking techniques.

1 Introduction

One of the most interesting and difficult problems in computer vision is that of recognizing and then tracking an entity of interest over time (i.e. in videos). This can be useful for surveillance of suspect individuals, analysis of video data (music videos, concerts, sports games, etc.), or in building robots that interact with their environment.

While tracking is a topic that has been extensively studied (see [1]), it is usually done so in isolation from recognition. To track the object of interest, the leading tracking algorithms mainly rely on properties of videos, notably that tracking target(s) will move by small amounts at a time. This allows the algorithms to detect/infer the object(s) positions from the previous ones by focusing on the region around the previously known position. While, these methods can work remarkably well on a clean video, such techniques have little hope when faced with long occlusions or frequent "frame switches" as these will cause a violation of the assumptions that the algorithms are based on.

Alternatively, object detection techniques could be applied frame by frame. This would result in a much more robust result since the algorithms make no assumptions concerning the relations between different frames. Hence, occlusions and frame switches pose no issue whatsoever. Unfortunately, even the best of these methods result in relatively noisy or inaccurate results when compared with the tracking algorithms. This can be attributed to the fact that the detection algorithms make no use of the wealth of information provided by the inter frame relationships (notably that most objects will usually not move by a large distance).

In this paper we propose a novel post-processing method, based on unsupervised clustering techniques, to "denoise" results obtained from object detection algorithms by exploiting inter-frame relationships. The result is a technique that is more robust to occlusions and frame switches than the standard tracking techniques and more accurate (less noisy) than the results obtained by naively applying object detection algorithms. We also have the added benefit of being able to easily do object recognition at the same time as detection. Our method will not only reduce the noise in the detection error (bad bounding boxes) but also in the classification error (bad labels).

2 Standard tracking methods and issues

Our problem, that of detecting objects throughout a video, or variations of it are often solved via tracking techniques. As mentioned above, the core idea is to exploit inter-frame relations, notably the assumption that most objects will not move by much between frames, to track or follow the various target objects. We quickly summarize the main tracking algorithms used in practice and show that they rely on core assumptions which are violated when faced with real life videos which contain frame switches, long occlusions, and various target objects coming in and out of the frames in variable number. In the presence of these realities, we show that these tracking algorithms are unusable.

As mentioned in [1], the most commonly used tracking algorithms are OLB [3],IVT [4], MIL [2], L1 [5], and TLD [6]. We give a quick summary of each and outline their main issue in our context.

[2] [3] [4] [5] [6]

3 Standard object detection and recognition methods

4 Our method: using clustering to exploit video structure in order to denoise detection and recognition methods

5 Experiments and Results

Acknowledgments

References

References

[1] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[2] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. Robust object tracking with online multiple instance learning. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1619–1632, 2011.

[3] Helmut Grabner, Michael Grabner, and Horst Bischof. Real-time tracking via on-line boosting. In *Bmvc*, volume 1, page 6, 2006.

[4] David A Ross, Jongwoo Lim, Ruei-Sung Lin, and Ming-Hsuan Yang. Incremental learning for robust visual tracking. *International journal of computer vision*, 77(1-3):125–141, 2008.

[5] Xue Mei and Haibin Ling. Robust visual tracking using l1 minimization. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1436–1443. IEEE, 2009.

[6] Zdenek Kalal, Jiri Matas, and Krystian Mikolajczyk. P-n learning: Bootstrapping binary classifiers by structural constraints. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 49–56. IEEE, 2010.