

Towards Scalable Analysis of Images and Videos

Abstract:

With widespread availability of low-cost devices capable of photo shooting and high-volume video recording, we are facing explosion of both image and video data. The sheer volume of such visual data poses both challenges and opportunities in machine learning and computer vision research.

In image classification, most of previous research has focused on small to medium-scale data sets, containing objects from dozens of categories. However, we could easily access images spreading thousands of categories. Unfortunately, despite the well-known advantages and recent advancements of multi-class classification techniques in machine learning, complexity concerns have driven most research on such super large-scale data set back to simple methods such as nearest neighbor search, one-vs-one or one-vs-rest approach. However, facing image classification problem with such huge task space, it is no surprise that these classical algorithms, often favored for their simplicity, will be brought to their knees not only because of the training time and storage cost they incur, but also because of the conceptual awkwardness of such algorithms in massive multi-class paradigms. Therefore, it is our goal to directly address the bigness of image data, not only the large number of training images and high-dimensional image features, but also the large task space. Specifically, we present algorithms capable of efficiently and effectively training classifiers that could differentiate tens of thousands of image classes.

Similar to images, one of the major difficulties in video analysis is also the huge amount of data, in the sense that videos could be hours long or even endless. However, it is often true that only a small portion of video contains important information. Consequently, algorithms that could automatically detect unusual events within streaming or archival video would significantly improve the efficiency of video analysis and save valuable human attention for only the most salient contents. Moreover, given lengthy recorded videos, such as those captured by digital cameras on mobile phones, or surveillance cameras, most users do not have the time or energy to edit the video such that only the most salient and interesting part of the original video is kept. To this end, we also develop algorithm for automatic video summarization. Finally, we further extend our research on video summarization into a supervised formulation, where users are asked to generate summaries for a subset of a class of videos of similar nature. Given such manually generated summaries, our algorithm learns the preferred storyline within the given class of videos, and automatically generates summaries for the rest of videos in the class, capturing the similar storyline as in those manually summarized videos.



Speaker:

Bin Zhao

PhD Candidate

Committee: Eric Xing, Chair
Alex Hauptmann
Tom Mitchell
Kristen Grauman (UT Austin)

Aug. 28, 2014

1:00pm

6501 GHC

Draft of the Thesis document:
http://www.cs.cmu.edu/~binzhao/thesis_draft.pdf

