Investigation of video analysis tools

Pre-Proposal in response to solicitation for Traffic21 Program Funding
Contact: Prof. Martial Hebert (hebert@ri.cmu.edu or 412-268-2585)

Project summary

We proposed to develop and evaluate video analysis tools for applications to data from transportation applications, both from infrastructure cameras and on-vehicle cameras. This topic is motivated by the rapid increase in the availability of video data and the fact that this data is largely under-utilized currently. Our main objectives are to use this seed project to 1) identify key sources of video data; 2) identify key features to recognize in videos for these applications; and 3) adapt our video algorithms and evaluate their performance. The proposed technical work will be based on our existing research on recognizing actions and behaviors in videos (primarily funded by NSF), using students who have already substantial experience in this research. We will have access to the necessary computing and storage facilities through our long-standing collaboration with the Intel Pittsburgh Lab.

The result of this seed project will be a set of proof-of-concept demonstrations, showing the operation of our video analysis algorithms in the context of real traffic-relevant data, both on fixed infrastructure and on vehicles. These results will then be used as support to approach potential sponsors with larger programs.

Problem statement and desired results

Overview

Video data is ubiquitous in transportation environments, e.g., from infrastructure equipment at crossings and overpasses, to cameras in transit vehicles. Automatic processing of this video data has the potential to generate critical information for traffic management and incident reporting which cannot be available through any other means. Opportunities for real-time processing include 1) Detection of anomalous behavior (of vehicles or pedestrians or riders) and 2) Understanding of traffic flow (of vehicles and/or pedestrians). In situations in which the data is recorded over long periods of time, off-line processing of the video data offers richer opportunities for estimating statistics relevant to traffic control, e.g., understanding pedestrian behavior at an intersection.

At the same time that the number of video sources increases exponentially, extensive research is under way in the computer vision field in the general area of video analysis. In particular, while earlier research focused exclusively on tracking objects in videos, recent advances have been in understanding the actions of the agents in the videos, e.g., recognizing specific actions from humans or recognizing specific behavior patterns. Under basic research funding from NSF and applied research funding from the Technology Collaborative, we have been working in this area for over six years. In particular, we have developed new techniques for recognizing actions and behaviors in videos[1].

[1] Yan Ke, Rahul Sukthankar, and Martial Hebert. Volumetric Features for Video Event Detection. International Journal of Computer Vision. 2009.

A representative example of our video analysis is shown in Figure 1: We recognize short-duration "events" or "actions" by comparing a configuration of shapes and motion learned from a small set of training images with data from an input video. Our approach can operate on videos of crowded scenes with a large number of actors moving in arbitrary ways, it does not rely on precise motion segmentation or feature extraction, and it is able to detect action patterns independently of the appearance of the agents. The technique handles multiple (possibly a large number of) training examples representing a single action, allowing us to better capture the variations across different instances of the event in the real-world. Figure 1(a) shows reference training videos and selected spatio-temporal templates displayed as a small set of 3D volumes in space-time. Figure 1(b) shows example outputs displaying the location at which the event is detected (shown inpink) in input videos.

We propose to use seed funding from the Traffic21 foundation to investigate the use of state-of-the-art computer vision tools to extract information from videos in the context of transportation and traffic applications. We believe that our group in particular in a unique position to bridge the gap between generic research in computer vision and concrete problems in transportation. In particular, the PI Prof. Hebert leads jointly a group of Ph.D. students conducting fundamental research in computer vision, and the Navlab group whose main mission is to apply machine perception to driving tasks. This synergy has been exploited in the past, for example in the development of state-of-the-art 3D machine perception algorithm for DoD unmanned ground vehicle. The spirit of this proposal is to use a small seed grant to bootstrap a similar effort in the area of video analysis. The expected outcome of this project is a set of results on real video data from traffic-related agencies which would enable to apply to larger efforts.

(a) Training (b) Detection

Figure 1. Example of video processing in which reference actions (shown in (a) in prototypical videos) are recognized in input videos (example frames in which the action is detected are shown in (b)).

Key technical challenges

Despite the recent successes, there are two major issues with deploying techniques from computer vision tools. First, they are generally developed and tested on generic datasets that are not representative of real applications, such as, for example, real data from infrastructure cameras. Second, the amount of data is typically quite limited and, in fact, most current developments in video analysis are data-starved. In contrast, existing installations, if recorded data is harvested, could generate unprecedented volumes of data, opening the door to new algorithms and new applications.

Pyry Matikainen, Martial Hebert, and Rahul Sukthankar. Trajectons: Action Recognition Through the Motion Analysis of Tracked Features. Workshop on Video-Oriented Object and Event Classification, ICCV 2009, September, 2009.

Pyry Matikainen, Martial Hebert, Rahul Sukthankar, and Yan Ke. Fast Motion Consistency Through Matrix Quantization. Proceedings of the British Machine Vision Conference, September, 2008, pp. 1055-1064.

Yan Ke, Rahul Sukthankar, and Martial Hebert. Event Detection in Crowded Videos. IEEE International Conference on Computer Vision, October, 2007.

Accordingly, we propose to use this seed project to:

1. Identify sources of video data from relevant agencies such as PAT, PennDoT. We expect to work closely with Traffic21 resources during this phase of the project.

2. Collect video data from the identified sources. Ideally, we would like to use two different sources corresponding to two substantially different types of environments and tasks, e.g., one from infrastructure cameras, the other from on-vehicle cameras. Presentation and publication of the results will not require the video data itself to be shown so that confidentiality agreement can be signed, if needed (e.g., guaranteeing anonymization of the faces in shown video clips).

3. Test and evaluate the performance of our algorithms on this data. We will first identify a small number of tasks, e.g., actions or motion features to be identified in the video, and then modify our algorithms based on the performance on this data set. Our final product will be a document summarizing the performance and the tasks that our algorithms were able to achieve and their key limitations.

One challenge in dealing with video data is the considerable amount of computing that is required, given the large amount of data. We plan to use our existing collaboration with the Intel Pittsburgh Lab (documented below) to supply the computing resources necessary for this type of work. We (including the students who be working on the project) have already extensive experience with this environment. The IP terms under which the background technology used in this project has been developed will allow its use in future programs.

Tasks and timetable

	1	2	3	4	5	6	7	8	9	10	11	12
Identification of sources of video data												
Data collection, conditioning and ground-truthing												
First evaluation based on existing research algorithms												
Algorithm refinement based on performance results												
Final evaluationand reporting												

Desired results

The desired outcome of this seed project is a set of proof-of-concept results demonstrating the operation of our video analysis algorithms in the context of real traffic-relevant data, both on fixed infrastructure and on vehicles. These results will then be used as support to approach potential sponsors with larger programs. This seed program would help bridge the gap between the need of the relevant agencies to use the massive amount of video data currently available and the current capabilities of computer vision tools. By taking advantage of the unique mix of competence available under one roof at the Robotics Institute, it would place CMU in a highly competitive position in this field. Beyond transportation, it would help in other efforts such as in the area of surveillance, for example.

Key personnel

Our team consists of a faculty member, Prof. Martial Hebert (PI), and one graduate student (Ph.D. in the Robotics Program) already working on video analysis algorithms and with extensive experience in experimenting with large video datasets, through the projects described above (NSF, TTC, collaboration with Intel). In addition, we expect to use undergraduate support for tasks such as labeling of the data, and implementation of basic evaluation tools.

Cost estimate

The costs are limited to the equivalent of the support of approximately half of a student time for one year. Existing computing resources will be used for this work so that no additional equipment or computing costs are required. This includes the software infrastructure for processing video data.

- Graduate student support: $30,000
- Undergraduate: $4,000
- Total: $34,000

Other partners

We have partnered with the Port Authority of Pittsburgh (PAT) in the past and we will use this connection to start our survey of data sources. We will expand the search to the other local and state agencies that manage video resources.

Over the past five years, we have collaborated closely with the Intel Pittsburgh Lab (Dr. Rahul Sukthankar) on the development of algorithms for video analysis. In addition to co-advising students and internship/fellowship opportunities, this collaboration gave us access to the Intel computing cluster and the computing infrastructure. This is a critical resource to support the processing of massive amounts of video data which is required in this type of research. We will continue to have full access to this computing resource for the proposed project, which will ensure that we can report results over a wide range of approaches and over a statistically significant amount of data.

Future R&D funding

We intend to apply for SBIR/STTR funding to commercialize the technology developed under this project. One potential small business partner is Mr. Bala Kumar. He is the co-founder of TimeSys Corporation (Pittsburgh), and served as the President of the company until 2001. TimeSys Corporation commercialized some of the DARPA SBIR work in Real-Time and Embedded Systems area. Since 2001, Mr. Kumar has been CEO of Etovia Systems (Pittsburgh). In addition, we intend to continue to pursue possible opportunities with Intel Pittsburgh.