

CIVIL ENGINEERING STUDIES
Illinois Center for Transportation Series No. 17-003
UIIU-ENG-2017-2003
ISSN: 0197-9191

OPPORTUNISTIC TRAFFIC SENSING USING EXISTING VIDEO SOURCES (PHASE II)

Prepared By
Jakob Eriksson
Yanzi Jin
Tomas Gerlich

University of Illinois at Chicago

Research Report No. FHWA-ICT-17-002

A report of the findings of
ICT PROJECT R27-169
Opportunistic Traffic Sensing
Using Existing Video Sources (Phase II)

Illinois Center for Transportation

February 2017

**ILLINOIS CENTER FOR
TRANSPORTATION**



TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-ICT-17-002		2. Government Accession No. N/A		3. Recipient's Catalog No. N/A	
4. Title and Subtitle Opportunistic Traffic Sensing Using Existing Video Sources (Phase II)				5. Report Date February 2017	
				6. Performing Organization Code N/A	
7. Author(s) Jakob Eriksson, Yanzi Jin, and Tomas Gerlich				8. Performing Organization Report No. ICT-17-003 UILU-ENG-2017-2003	
9. Performing Organization Name and Address Department of Computer Science College of Engineering University of Illinois at Chicago Chicago, IL 60607				10. Work Unit No. N/A	
				11. Contract or Grant No. R27-169	
12. Sponsoring Agency Name and Address Illinois Department of Transportation (SPR) Bureau of Materials and Physical Research 126 East Ash Street Springfield, IL 62704				13. Type of Report and Period Covered May 16, 2015–February 15, 2017	
				14. Sponsoring Agency Code FHWA	
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.					
16. Abstract <p>The purpose of the project reported on here was to investigate methods for automatic traffic sensing using traffic surveillance cameras, red light cameras, and other permanent and pre-existing video sources. Success in this direction would potentially yield the ability to produce continuous, daily traffic counts where such video sources exist, as compared to the occasional traffic studies performed today.</p> <p>The methods investigated come from the field of computer vision, including optical flow, background subtraction, and object detection and tracking, as well as control theory for the fusing of the results of these methods. Our system outperforms the state of the art in vehicle tracking, and it runs at faster frame rate. More work remains in improving robustness to occlusion and to improve accuracy of nighttime imagery. Our work on rigid motion optical flow was published in the proceedings of the International Conference on 3D Vision, and our work on vehicle tracking is currently under submission to the IEEE Winter Conference on Applications of Computer Vision.</p>					
17. Key Words computer vision, vehicle tracking, optical flow			18. Distribution Statement No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified.		20. Security Classif. (of this page) Unclassified.		21. No. of Pages 11 pp.	22. Price N/A

ACKNOWLEDGMENT, DISCLAIMER, MANUFACTURERS' NAMES

This publication is based on the results of **ICT-R27-169, Opportunistic Traffic Sensing Using Existing Video Sources (Phase II)**. ICT-R27-169 was conducted in cooperation with the Illinois Center for Transportation; the Illinois Department of Transportation; and the U.S. Department of Transportation, Federal Highway Administration.

Members of the Technical Review panel were the following:

- William Morgan, IDOT, TRP Chair
- Vince Durante, IDOT
- Mike Miller, IDOT
- David Pulsipher, City of Chicago

The contents of this report reflect the view of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Illinois Center for Transportation, the Illinois Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

Trademark or manufacturers' names appear in this report only because they are considered essential to the object of this document and do not constitute an endorsement of product by the Federal Highway Administration, the Illinois Department of Transportation, or the Illinois Center for Transportation.

EXECUTIVE SUMMARY

The purpose of the project reported on here was to investigate methods for automatic traffic sensing using traffic surveillance cameras, red light cameras, and other permanent and pre-existing video sources. Success in this direction would potentially yield the ability to produce continuous, daily traffic counts where such video sources exist, as compared to the occasional traffic studies performed today.

Analyzing video from existing sources differs significantly from analyzing video collected for the purpose of traffic analysis. In particular, purpose-collected video typically has a high degree of control over perspective, coverage, weather, image quality, and lighting, whereas existing video cannot be made to fit any such constraints. Some constraints can be met by selection, such as choosing to analyze only summertime videos recorded during daylight hours, by high-quality cameras that offer a favorable perspective. However, to maximize the utility and applicability of a system meant for existing video sources, the system must support a wide range of challenging conditions.

The methods investigated come from the field of computer vision, including optical flow, background subtraction, and object detection and tracking, as well as control theory for the fusing of the results of these methods. While the focus has been on applying existing methods to a new problem, we have made significant contributions to the literature on optical flow and object tracking and have proposed a new information fusion method for combining the information gleaned from a variety of sources into a coherent final result.

To evaluate our work, we painstakingly collected a set of video clips with associated ground-truth annotations. These annotations show, for each individual video frame, the size and location of each vehicle present in the scene, as well as its movement between frames. Using this dataset, we were able to produce detailed evaluation results for our own methods and those of others, which guided the development of our system.

Using our ground-truth dataset for comparative evaluation, we determined that our vehicle tracking system outperforms the state of the art in object tracking in terms of accuracy. It also adds automatic handling of scene entry and exit and runs five times faster. That said, more research is needed, primarily to improve robustness to occlusion and to improve accuracy on nighttime imagery. Our work on rigid-motion optical flow was published in the proceedings of the International Conference on 3D Vision, and our work on vehicle tracking is currently under submission to the IEEE Winter Conference on Applications of Computer Vision.

CONTENTS

- CHAPTER 1: BACKGROUND AND MOTIVATION.....1
 - 1.1 EXISTING VIDEO SOURCES1
 - 1.2 APPLICATIONS OF AUTOMATIC VEHICLE TRACKING IN OPPORTUNISTIC VIDEO.....1
 - 1.3 PRIMARY CHALLENGES IN COMPUTER VISION–BASED TRAFFIC MEASUREMENTS
FROM EXISTING VIDEO SOURCES.....1
- CHAPTER 2: VEHICLE TRACKER DESIGN.....2
- CHAPTER 3: VEHICLE TRACKER EVALUATION4
- CHAPTER 4: OPTICAL FLOW FOR RIGID MULTI-MOTION SCENES.....7
- CHAPTER 5: USER INTERFACE10
- CHAPTER 6: NEXT STEPS.....11

CHAPTER 1: BACKGROUND AND MOTIVATION

1.1 EXISTING VIDEO SOURCES

The state of Illinois and the city of Chicago operate extensive networks of video cameras facing roadways for a variety of purposes, including traffic monitoring, emergency response, and law enforcement. Many of these cameras are accessible remotely, which enables the collection of video on demand, and at little or no cost.

In principle, these video resources can already be used for traffic analysis. However, the process is extremely labor intensive because it requires a person to manually count each vehicle captured by the video recording. This project is based on the hypothesis that this process could be fully, or at least largely, automated.

1.2 APPLICATIONS OF AUTOMATIC VEHICLE TRACKING IN OPPORTUNISTIC VIDEO

Having the ability to analyze video from existing sources to produce traffic information enables a variety of uses that is impossible or impractical today. Ideally, a fully automatic 24/7 analysis system would enable continuous monitoring of traffic conditions on most major thoroughfares of the state, both at real time and for purposes of historical analysis.

Barring such a widespread and widely applicable implementation, the proposed system could be used for impromptu traffic studies, to quickly measure the impact of changes to traffic patterns or signaling, or to produce ADT counts without the relatively extensive preparation, cost, and analysis time required by current traffic count services.

1.3 PRIMARY CHALLENGES IN COMPUTER VISION–BASED TRAFFIC MEASUREMENTS FROM EXISTING VIDEO SOURCES

Compared to special-purpose video recordings, using existing/opportunistic video implies a wide range of quality, perspective, and lighting. This in turn creates challenges in vehicle detection, tracking vehicles through occlusion and scale changes, and avoiding spurious detections.

CHAPTER 2: VEHICLE TRACKER DESIGN

Figure 1 illustrates the overall design of our vehicle tracking system. For each frame of the video, a preliminary computation step generates *boxes*, which indicate the location and extent of potential vehicles in the scene, as well as optical flow, which indicates the direction and magnitude of movement for each individual pixel. Here, optical flow is computed with respect to the previous frame.

The boxes are produced by two separate processes: One process is based on a standard object detector, which scans the image for image patches that look similar to vehicles. This system works well for well-lit, high-resolution imagery, but it usually does not produce many detections when the image quality is poor. The other process, background subtraction, maintains a continuously updated model of what the background of the scene looks like. When a vehicle enters the scene, its appearance typically differs from the background and is detected by the background subtraction process, which typically produces a foreground box approximately encompassing the vehicle. Background subtraction typically detects vehicles under a wide variety of conditions. However, it has a tendency to produce spurious detections, and it often fails to detect a vehicle after it has been stationary for some time.

The optical flow produces a flow field consisting of a two-dimensional vector for each pixel of the frame. Here, the process of producing a high-quality optical flow field is a research area of its own, and our system can work with any optical flow algorithm. We have also developed our own optical flow algorithm for rigid motion scenes, which was published in the proceedings of the International Conference of 3D Vision in 2016. Optical flow itself is not used to detect vehicles—it is highly error prone and often quite sparse for uniform-colored objects, and it does not lend itself to detection. However, given a set of boxes, optical flow can be very helpful in estimating motion.

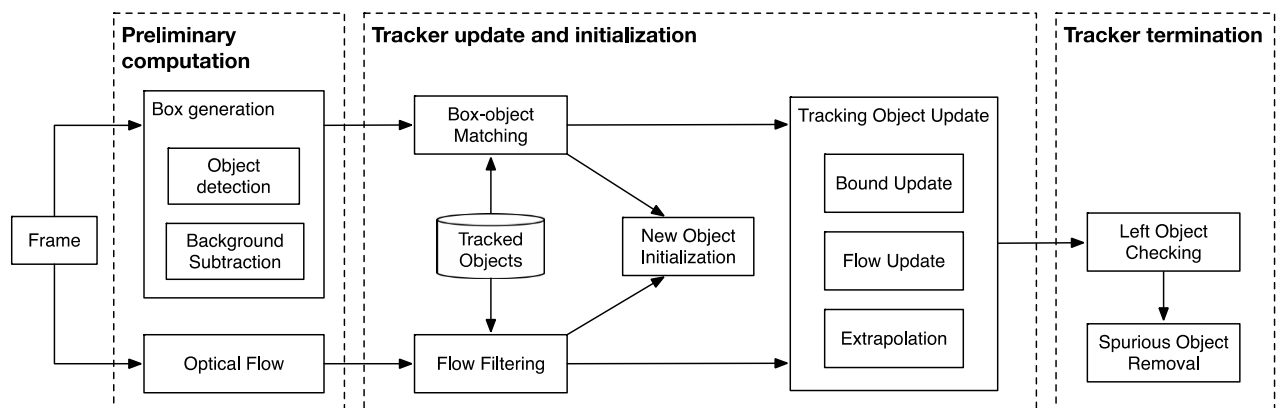


Figure 1. High-level design of our current vehicle tracking system. Counting is done in post-processing, by matching vehicle tracks against user-provided templates.

Following the preliminary computation step is tracker update and initialization. Here, the boxes and flow produced in preliminary computation are combined with a current set of tracked objects, both to detect the appearance of new objects and to update the size and location of each currently tracked object. In addition to tracking size and location, each object maintains an internal state consisting of size, location, speed, acceleration, and rate of size change. By tracking hidden variables such as speed and acceleration, our system is able to predict the future location of a vehicle, to better match a tracked object with its corresponding box in a new frame.

CHAPTER 3: VEHICLE TRACKER EVALUATION

To fully evaluate the performance of a vehicle tracking system, we need a collection of videos annotated with the “ground truth” size and location and movements of vehicles present in the scene. Figure 2 illustrates the dataset that we have created. It consists of 13 videos, each 5 minutes long, with the size, location, and movement of every vehicle painstakingly annotated. Figure 3 shows a screen shot of the annotation system in use.

Table 2: Dataset overview. The last five columns indicates the number of videos containing the corresponding scenario.

Video group	Resolution	Object size (pixels)	Total number	Night	Occlusion	Shadow	Distortion	Pedestrian
lowRes	342 × 228	32-44814	5	0	3	1	3	0
	320 × 240	48-25284	2	0	2	1	2	0
highRes	720 × 576	84-255106	3	0	3	0	0	1
night	640 × 480	81-134520	3	2	3	0	0	2



Figure 6: Snapshots of videos in our dataset, with various resolution, viewpoint, illumination condition, vehicle size and interactions. We group these videos by their characteristics: lowRes, highRes and night. Top row is simple low resolution videos, (h) and (i) are night videos, the rest (j) - (n) are complex high resolution videos. Note that the last two (m) and (n) are from the same video.

Figure 2. Ground-truth dataset overview.

Using this ground-truth dataset, we are able to produce a highly reliable, detailed, and quantitative evaluation of various aspects of our system, as well as the final tracking output. This is essential in guiding the development of the tracking system, as well as in comparing the performance of our system to previous work in this area.

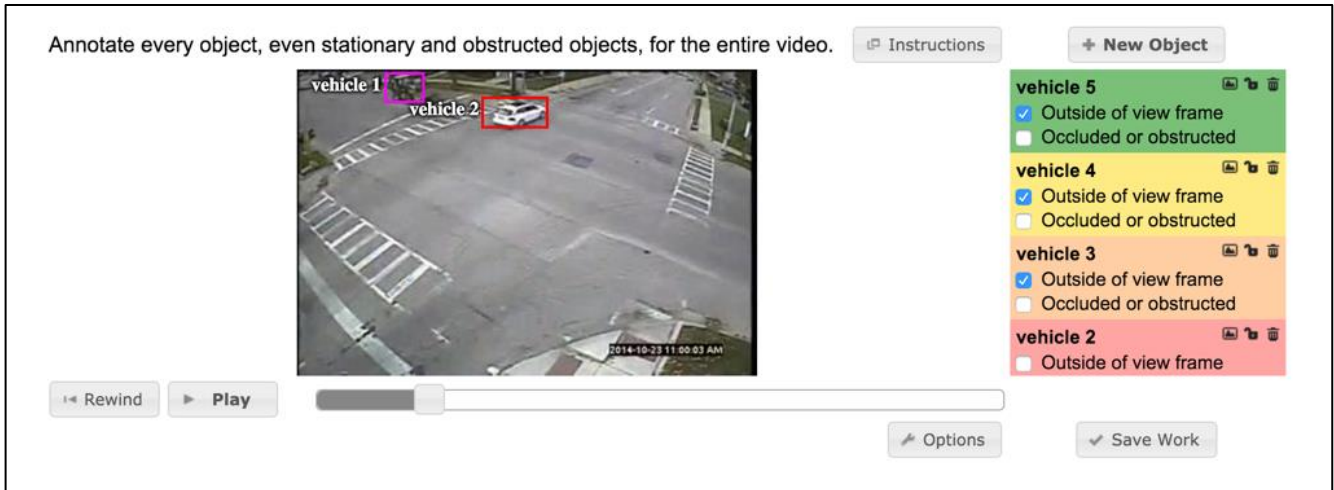


Figure 3. Screenshot of the ground-truth annotation system in use.

One basic evaluation measure is the number of objects detected by the system, as well as a breakdown of objects that were actual vehicles vs. other objects or spurious detections. Figure 4 illustrates the counting performance of our system on our evaluation dataset. Here, true positives are objects that matched with an object in the ground truth, and false positives are objects reported that did not match with the ground truth. In most cases, we found that the false positives reported were due to double-counting, either where a single vehicle was detected as two pieces, or where we first lost track of a vehicle then rediscovered it and reported it as a second vehicle. We report results for three different types of trackers: BG, which uses only background subtraction for box generation; DET, which uses only the object detector; and BG+DET, which combines the two.

Overall, we find that the combined system provides acceptably accurate counts on most videos, but it underperforms on night videos. We also find that complex videos, where occlusion is common and the scene tends to be crowded, pose a greater challenge to our system than poor quality video does.

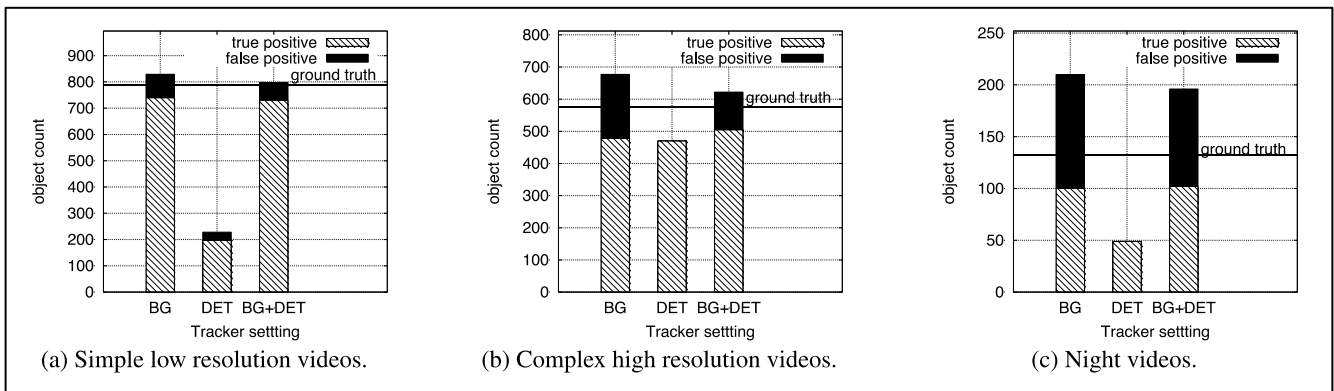


Figure 4. Object count accuracy vs. ground truth.

We also evaluated the system’s performance against the state of the art in object tracking. Here, because no existing systems were available for end-to-end vehicle tracking and counting, we compare only against a dedicated, state-of-the-art object tracker. This tracker (STRUCK) does not provide automatic initialization, which we instead provide from our ground-truth annotations. Thus, a direct comparison is not quite fair because STRUCK is not a practical solution. That said, Figure 5 illustrates the tracking performance of our system vs. STRUCK, where a higher overlap ratio is better. The horizontal lines show performance with automatic initialization enabled, and the curves show performance for various initialization thresholds: these decide when to initialize tracking during the object’s lifetime. Later initialization results in lost tracking initially, whereas early initialization tends to produce lower quality tracking. Overall, we find that our system substantially outperforms the state of the art in all except night videos.

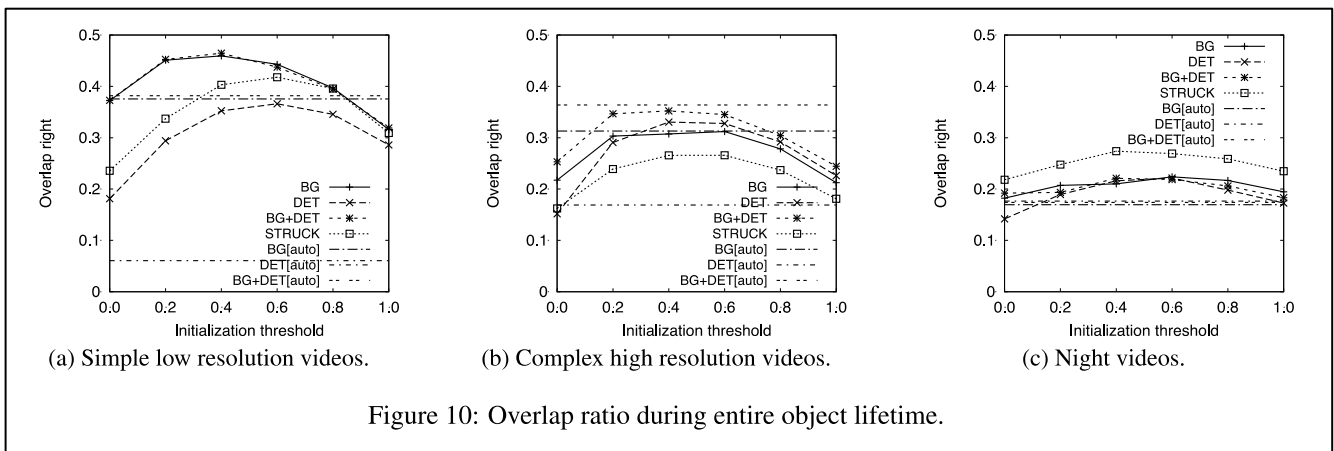


Figure 10: Overlap ratio during entire object lifetime.

Figure 5. Tracking accuracy vs. state of the art.

CHAPTER 4: OPTICAL FLOW FOR RIGID MULTI-MOTION SCENES

Based on our experiments with optical flow algorithms from the literature, we discovered an opportunity for accuracy improvement in traffic scenes. In general, optical flow is a highly underconstrained problem, meaning there are potentially very many optical flow fields that explain the apparent changes between video frames. To select one of these very many solutions, generic optical flow algorithms usually introduce a basic “smoothness” assumption, saying that the flow in adjacent pixels tends to change smoothly. However, we have significant additional knowledge about the movement in our scenes. Specifically, we know that the objects of interest are largely rigid, and they move in certain highly constrained ways—essentially moving only in straight lines and turns. By introducing these additional and novel constraints, we were able to significantly improve on the performance of state-of-the art algorithms for generic optical flow.

To quantitatively evaluate the accuracy of our system, we created a dataset consisting of synthetic but photo-realistic imagery of several traffic scenes, generated using computer rendering software. In addition to producing the imagery, we modified the software to also output the exact optical flow ground truth for the scene. We then computed the optical flow for a pair of frames using our method as well as several other methods from the literature. Figure 6 lists the results, with **MMSGM-fGT** and **MMSGM-fGT-EF** representing our system. Here, the bold lettering indicates the best-performing system for each type of flow error, with our system matching or substantially outperforming the state of the art in most categories. Figure 7 illustrates these results qualitatively for several example images, with our system displaying a significant advantage.

Method	fg (%)	fg (px)	bg (%)	bg (px)	comb (%)	comb (px)
1280x720						
MMSGM-fGT	26.4	8.87	2.30	1.80	6.96	3.45
MMSGM-fGT+EF	26.1	6.91	1.23	0.46	5.92	1.89
DISCFLOW+EF [11]	36.3	8.41	1.53	0.44	8.08	2.14
DM+EF [13]	49.4	12.7	2.26	0.42	11.2	3.05
FLOWNET [5]	71.3	13.5	10.6	2.31	22.1	4.59
CNL [15]	68.1	29.4	13.0	6.12	22.8	10.7
320x180 (QR)						
MMSGM-fGT	14.8	2.91	4.92	1.09	7.11	1.54
MMSGM-fGT+EF	21.3	3.06	2.48	0.33	5.76	0.79
MMSGM-f64	19.7	3.77	5.97	1.35	8.90	1.93
MMSGM-f64+EF	18.8	2.35	3.71	0.41	6.79	0.83
DISCFLOW+EF	22.6	3.26	2.66	0.39	6.30	0.89
DM+EF	29.1	3.40	1.26	0.18	6.79	0.86
FLOWNET	76.0	7.64	17.3	1.95	27.0	3.05
CNL	46.2	7.57	5.71	0.77	13.5	2.20

Table 1. **Top:** algorithms were tested on full resolution *FlowTruth* testing dataset where both the % outliers above 3px threshold measure and the end-point error (px) for non-occluded pixels w.r.t ground truth is averaged over the 6 photo-realistic scenes. **Bottom:** same as above with input images resized to quarter size.

Figure 6. Quantitative results: Rigid optical flow accuracy vs. several optical flow competitors.



Figure 2. Qualitative evaluation on our novel synthetic dataset. QR denotes quarter-resolution imagery: f64 was not practical to run on the full resolution images due to memory constraints.

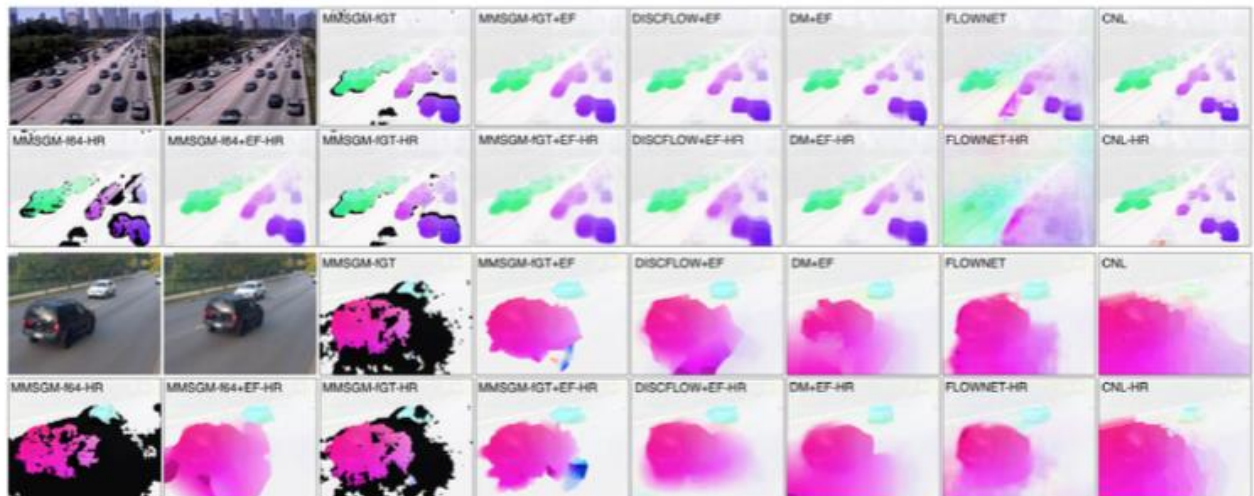


Figure 3. Qualitative evaluation on real-world traffic-camera imagery (top 648x500, bottom 720x576). HR denotes half-resolution.

Figure 7. Qualitative results: Rigid optical flow accuracy vs. several optical flow competitors.

CHAPTER 5: USER INTERFACE

One of the goals of this project was to create a means by which IDOT staff can directly use the system. Given our existing software, we decided to pursue a virtual desktop approach to create this facility. Here, the user interacts remotely with an installation on the existing UIC computer infrastructure. On the user side, two pieces of standard software are installed: scp for transferring video files to UIC, and VNC for establishing a remote desktop connection.

Figure 8 shows a screen shot of the application in operation. Here, the user indicates several motion templates—that is, movements that the user is interested in counting. The system then processes the video and outputs a number for each motion template, indicating the number of vehicles that followed that template throughout the video.

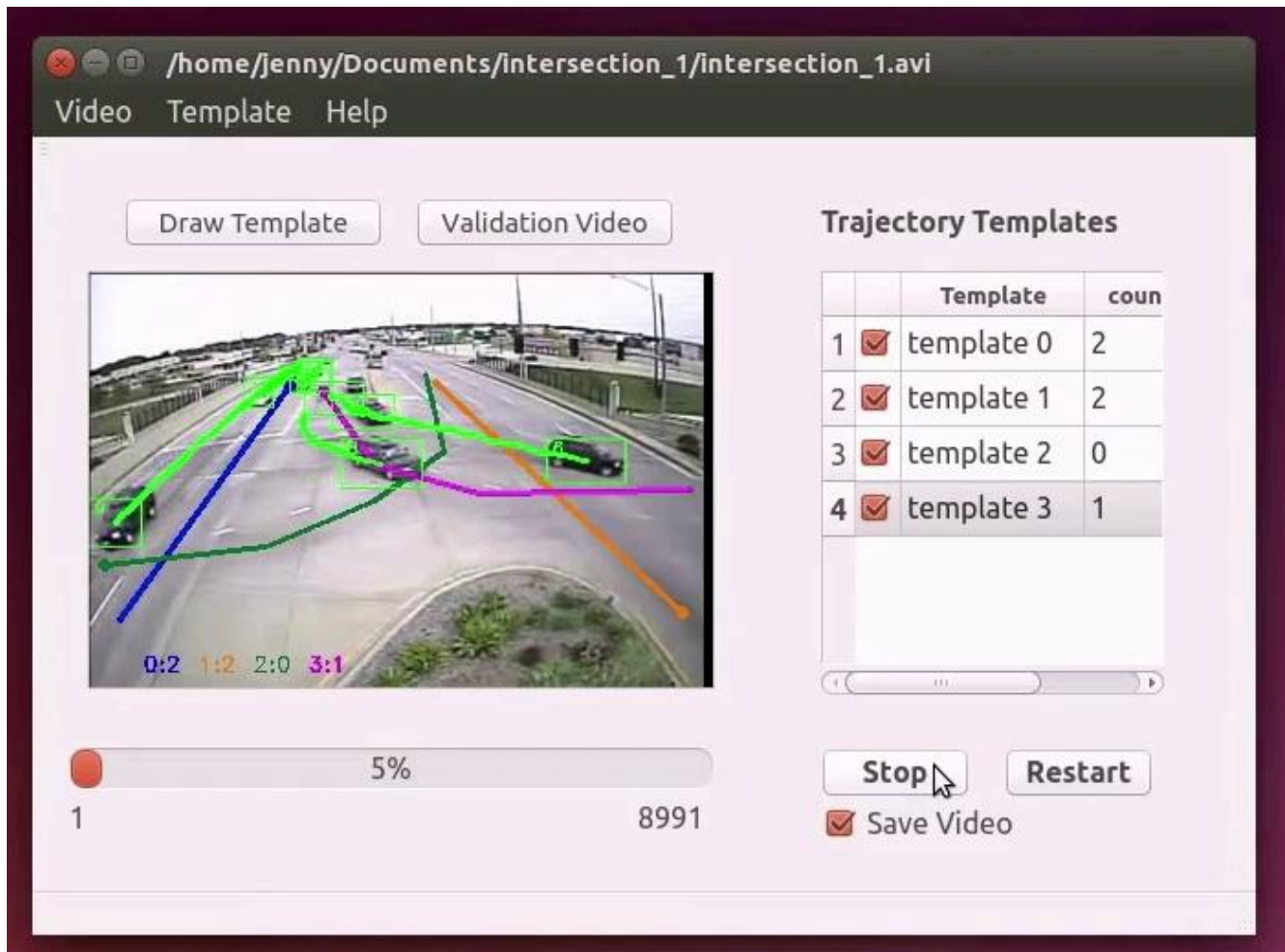


Figure 8. Screen shot of vehicle-counting application in use.

CHAPTER 6: NEXT STEPS

We plan to submit a proposal for a continuation of this project, with the goal of improving tracking and counting accuracy and widening applicability to more challenging scenes and conditions. Moreover, while the remote desktop solution for the user interface was practical from a development effort point of view, the user experience was not ideal. We plan to propose a Web-based solution, using an identical processing pipeline but allowing users to upload videos and manage counting and other processing using a revamped and entirely Web-based interface. We expect this to dramatically improve the user experience, which is an important yet currently underdeveloped aspect of the system.

