

Motion Classification using a Linear Dynamical System

ECE 199 Research Proposal - Summer 2007

Mulloy Morrow, Student Researcher
Antoni Chan, Advisor
Professor Nuno Vasconcelos, Advisor
Statistical Visual Computing Laboratory
University of California, San Diego

“Abstract – We propose to model the traffic flow in a video using a holistic generative model that does not require segmentation or tracking. In particular, we adopt the dynamic texture model, an autoregressive stochastic process, which encodes the appearance and the underlying motion separately into two probability distributions. With this representation, retrieval of similar video sequences and classification of traffic congestion can be performed using the Kullback-Leibler divergence and the Martin distance. Experimental results show good retrieval and classification performance, with robustness to environmental conditions such as variable lighting and shadows.”ⁱ

As mention above, we are investigating the utility of a linear dynamic system as the input space for a few motion classification techniques. In this approach, the spatial appearance and underlying motion of video is modeled separately. However, these models are intrinsically constructed from the three dimensions of video (interframe data).

The first two dimensions are what is shown in any one frame. The third dimension is time. Older methods of motion classification involved dealing with individual frames independently (i.e. segmentation) and tracking recognized shapes through an array of frames. Our classification approach intrinsically includes all three dimensions in feature extraction. Therefore, classification may be performed on video with interframe language. A necessity in motion classification of video containing only fluidity.

After our linear dynamic system is constructed, “classification is performed by selecting the category corresponding to the Gaussian-mixture hidden Markov model (GM-HMM) of largest likelihood for the query video.”ⁱⁱ This has been experimentally shown to be an accurate process for congestion levels on a freeway. However, we’re interested in discovering the latitude in classifiers this method promises.

Our objective is to implement our method of classification on many interesting scenes. On page 2 is a list of scenes filmed. We have compiled a video database with over 24 hours of video of over 10 different scenes. Each scene’s subject is fundamentally either pedestrian or vehicular traffic flow.

Our goal for each scene was to film from three different angles, each angle having both a panning and still shot. Scenes were filmed on multiple occasions at different times of day in order to capture different levels of flow. The goal of each scene was to build simple motion classifiers that would distinguish between congestion levels and types of traffic flow (i.e. pedestrians exiting a building or cars turning left at an intersection).

We will demonstrate the robust classification of pedestrian as well as vehicular traffic flow scenes (i.e. high, medium, or low congestion, entering, exiting, walking away from us or towards us).

Currently the video database we have compiled is on the web at:

<www.svcl.ucsd.edu/projects/motiondb>

Scenes Shot

Pedestrian Throughway

01. Sidewalk 01 (near CG BLFG)
02. Sidewalk 02 (near Economics BLDG)
03. UCSD Library Walk
04. Sidewalk 03 (Downtown Fifth Ave.)
05. Parade (Hillcrest)
06. Plaza (UCSD Center Hall)

Pedestrian Entrance/Exit

07. Geisel Library
08. Sea World Exit
09. Atkinson Hall (Emergency Evacuation)
10. Concert (Not shot yet)

Vehicular Throughway

11. Intersection 01
12. Intersection 02

Vehicular Entrance/Exit

13. Freeway Entrance
14. Parking Structure

Process Overview

Videography:

Video shot using Sony DCR-HC38 HandyCam.

All angles were shot with camera zoom fully wide and tripod legs fully extended unless otherwise noted. However, most panning shots were shot with center arm not extended for better stability in manually rotating camera.

Care was taken to shoot each angle completely before moving tripod to another location. However, not all traffic scenes are capable of being shot on the same day. Therefore, some slight variation in an angle's view was incurred. This error was minimized either by (1) cropping the video so that we had the same view in frame across the board, or (2) re-shooting the entire scene over for more consistency.

After shooting, video is cropped temporally to cut out outliers and obtain a pure class (i.e. just low traffic or just cars turning right). Then video dimensions were reduced to 50% and compressed using codecs: MPEG-4 (.mp4), H.264 (.mov), cinepak (.avi). Care was taken to not compress excessively as noticeable artifacts would occur.

Directories:

- 1) < /data2/vidtex/mulloy > - *Contains three main folders.*
 - a) < ../video > - *Contains raw video which as filmed.*
 - b) < ../Data > - *Contains video clips that went through 'vid2y48.m'*
 - c) < ../MetaData > - *Contains:*
 - i) run_experiment.m - *Executes script to model processed video clips, to train classifiers, to test classifiers, to validate classifiers, to see error results. Experiment parameters are set for particular scene.*
 - ii) ImageMaster.mat - *Contains list of pre-processed video directories and their respective class (e.g. high, medium, low)*
 - iii) exp**.mat - *Contains workspace from executing 'run_experiment.m' in Matlab. Contains experimental results.*

- 2) ~/vidtex - *Contains matlab m-files used by programs such as 'vid2y48.m' and 'run_experiment.m'*

Programs:

- 1) Matlab
- 2) vid2y48.m -- *m-file within matlab. This program must be ran on server "weiner" as this server has the codecs for videos.*
- 3) run_experiment.m --

Process:

1. Run vid2y48.m in Matlab. - *This program converts full length color clips into arbitrarily long B&W data clips (usually 50 frames). This program also outputs*

'ImageMaster' which is a metadata file containing the directory and class associated with every output clip.

2. Run run_experiment.m - *located in each MetaData file with scene specific parameters.*

3. Based on results, adjust experiment parameters.

4. Repeat steps 2 and 3.

01. Sidewalk01

(Pedestrian Throughway)

Shot Description: Shot near Cognitive Science Building and singing tree. **Three angles:** (1) Oblique - 6-foot tripod fully extended above heads and placed on boulder to give slight birds-eye-view, facing east towards library walk. (2) Front - eye level, two tripod legs on crack in cement facing west, (3) Side - eye-level in dirt facing north.



Filming Status:

Still Shots -

- (1) Oblique Angle - *Missing* 'medium' traffic level. However, can make this up using a 'medium-high' class.
- (2) Front Angle - *Complete*. Two slightly different versions.
- (3) Side Angle - *Missing*: 'high' level. Can't shoot this level until october.

Panning Shots -

No Panning shots were filmed due to time constraints.

Traffic-Level Classifier Description: Discriminating between *high*, *medium*, and *low* flows of pedestrian traffic. Three experiments, one per angle.

(1) *Oblique* - Initial Experiment Results:

Although a distinct 'medium' class was not shot yet, we were able to achieve high accuracy using two levels of high. (i.e. 'low-high' & 'high-high').

md_error_mean	=	0.1659	=	83.41%	Accurate
kl_error_mean	=	0.2505	=	74.95%	
kl_svmerror_mean	=	0.3141	=	68.59%	
ikl_error_mean	=	0.3676	=	63.24%	
ikl_svmerror_mean	=	0.3004	=	69.96%	

(2) *Front* - Initial Experiment Results:

Waiting on results. ETA: less than one week.

(3) *Side* - Initial Experiment Results:

Waiting on 'high' class. ETA: October.

02. Sidewalk02

(Pedestrian Throughway)

Shot Description: Shot near Economics Building on Ridgewalk, mid-day. **Three angles:** (1) Front - *eye level, set up under bridge, facing south*, (2) Side - *eye-level facing west*. (3) Oblique - *Shot on bridge, in middle, facing south*. 'Medium' and 'low' are readily available. However, 'Heavy' was captured during passing periods.



Filming Status:

Still Shots -

(1) Oblique Angle - *Complete*.

(2) Front Angle - *Complete*.

(3) Side Angle - Missing 'high' level. Can't shoot this level until October.

Panning Shots -

Panning-Low-Front was shot successfully.

Traffic-Level Classifier Description: Discriminating between *high*, *medium*, and *low* flows of pedestrian traffic. Three experiments, one per angle.

(1) **Front - Initial Experiment Results:**

(2) **Side - Initial Experiment Results:**

(3) **Oblique - Initial Experiment Results:**

03. Library Walk

(Pedestrian Throughway)

Shot Description: Library Walk is a throughway used by thousands throughout the day. At times there are so many people using this throughway that it is difficult to even walk.

Three angles: (1) Front - *eye level, set up in center of walkway facing south*, (2) Side - *eye-level facing west*. (3) Oblique - *Shot from center of 5th floor of Geisel Library, UCSD, mid-day and facing south*.



Filming Status:

Still Shots -

- (1) Oblique Angle - *Complete*. Two versions of this angle. Shot at different times of year.
- (2) Front Angle - *Complete*.
- (3) Side Angle - Missing '*high*' and '*medium*' levels. Can't shoot this level until october.

Panning Shots -

Missing Panning-Front.

Panning Oblique is not compatible.

Traffic-Level Classifier Description: Discriminating between *high*, *medium*, and *low* flows of pedestrian traffic. Three experiments, one per angle.

(1) *Front* - Initial Experiment Results:

md_error_mean = 0.3558 = 64% Accurate
 kl_error_mean = 0.3657 = 64%
 kl_svmerror_mean = 0.4138 = 58%
 ikl_error_mean = 0.3800 = 62%
 ikl_svmerror_mean = 0.2983 = 70%

(2) *Side* - Initial Experiment Results:

Waiting on '*high*' and '*medium*' classes.

(3) *Oblique* - Initial Experiment Results:

md_error_mean = 0.4314

kl_error_mean = 0.5329
kl_svmerror_mean = 0.4155
ikl_error_mean = 0.5611
ikl_svmerror_mean = 0.4546

(3) *Oblique* Expectations: These were the errors before the data was separated into two versions. Better accuracy is expected from each version after this distinction is ran through experiment. However, these results show that there is a large error associated with calibration.

04. Sidewalk03

(Pedestrian Throughway)

Shot Description: Fifth Ave in Downtown San Diego's Gas Lamp District is a tourist hot spot. Pedestrians are highest in the evening. This area provides a very interesting space to study. It contains a close interaction between pedestrians, people waiting for buses, people walking in and out of shops and restaurants, and vehicular traffic. First

Three angles shot on east side of street in front of Urban Outfitters: **(1)** Front - *eye level, set up in center of walkway facing south*, **(2)** Side - *eye-level facing east*. **(3)** Oblique - *Shot from curb, facing sidewalk from oblique angle*.

Filming Status:

Still Shots -

(1) Oblique Angle -

(2) Front Angle -

(3) Side Angle -

Panning Shots -

Traffic-Level Classifier Description: Discriminating between *high, medium*, and *low* flows of pedestrian traffic. Three experiments, one per angle.

(1) Front - Initial Experiment Results:

(2) Side - Initial Experiment Results:

(3) Oblique - Initial Experiment Results:

05. Parade

(Pedestrian Throughway)

Shot Description: Pride Parade Summer 2006. Shot the streets just after the parade ended. The streets were taken over by pedestrians. Most walking in the same direction.

Filmed Three Angles plus walking footage. Three Angles: **(1)** Front - eye level facing pedestrians as they walked towards the camera, **(2)** Side - eye level facing perpendicular to the motion of the crowd., **(3)** Oblique - off at an angle between the first two angles, above eye level. **Walking:** walked with crowd, camera at chest level.

06. Plaza

(Pedestrian Throughway)

Shot Description: Center Hall Plaza, at the end of library walk, south of the library, is a large area that many use between and during classes. An interesting space that contains pedestrians, bikers, people sitting, people talking but not walking, and campus tours. **One angle:** (1) Oblique - *second story oblique angle of entire court yard*. Expected angles: (2) Front - *eye level on library walk facing south*. (3) Side - *side of library walk, facing south-east*.



Filming Status:

Still Shots -

- (1) Oblique Angle - *Complete*
- (2) Front Angle - *Not Shot*
- (3) Side Angle - *Not Shot*

Panning Shots -

Traffic-Level Classifier Description: Discriminating between *high*, *medium*, and *low* flows of pedestrian traffic. Three experiments, one per angle.

(1) *Oblique* - Initial Experiment Results:

md_error_mean	=	0.3259 = 67%
kl_error_mean	=	0.3111 = 68%
kl_svmerror_mean	=	0.2733 = 72%
ikl_error_mean	=	0.2734 = 72%
ikl_svmerror_mean	=	0.2810 = 72%

07. Geisel Library

(Pedestrian Entrance /Exit)

Shot Description: Geisel Library, located near Price Center on campus at the University of California, San Diego. The Library has many floors all filled with books and computers. A steady flow of traffic can be seen constantly at the entrance, even during summer sessions.

Three Angles: (1) Front - eye level, set up was on the edge of a semi-circle of concrete surrounding a few trees. Back two legs of tripod were lined up with joint of two cement segments. This is left of the silent tree, looking north. (2) Side - eye-level facing east. Set up was on inside of library in the 'Seuss Room,' shooting out through windows. (3) Oblique - Shot from center of 5th floor of Geisel Library, UCSD, mid-day and facing south. Revealing a birds-eye-view of the walkway just outside the library main entrance.



Filming Status:

Still Shots -

- (1) Front Angle - Shoot complete, needs processing.
- (2) Side Angle - Not enough traffic variation in tight angle.
- (3) Oblique Angle - *Complete*

Panning Shots -

- (1) Front Angle - Not shot
- (2) Side Angle - *Complete*
- (3) Oblique - *Complete*

Traffic-Level Classifier Description: Discriminating between *high*, *medium*, and *low* flows of pedestrian traffic. Three experiments, one per angle.

- (1) **Front - Initial Experiment Results:**
- (2) **Side - Initial Experiment Results:**
- (3) **Oblique - Initial Experiment Results:**

Entering vs Exiting Classifier Description: Discriminating between pedestrians *Entering*, *Exiting*, and a *Balance* of the two. Three experiments, one per angle.

- 1) **Front - Initial Experiment Results:**

(2) *Side* - Initial Experiment Results:

(3) *Oblique* - Initial Experiment Results:

08. Sea World Exit

(Pedestrian Exit)

Shot Description: Shot at the exit from two different angles. **(1)** Front - *eye level and head on facing the people exiting*, **(2)** Side - *eye level and facing perpendicular to motion*.

Filming Status: Incomplete

Expectations: Will not use footage.

09. Atkinson Hall

(Pedestrian Entrance/Exit)

Shot Description:

“Approximately 250 law enforcement and public safety personnel went into action at the home of the UCSD division of Calit2, including a rappel from a SWAT helicopter onto the roof.

The San Diego Metropolitan Medical Strike Team (MMST) drill, named Operation College Freedom, was an opportunity for Calit2 researchers to demonstrate and evaluate new technologies which they are developing for the management of disaster and mass-casualty situations.”ⁱⁱⁱ Unfortunately, there was not a massive outpour of victims as was expected.

Four Angles: **(1)** Front - *eye level, set up outside of entrance, facing north*, **(2)** Side - *eye-level facing east, on side of building where all victims passed by new triage sensors*. **(3)** Oblique - *Shot from .roof top of theater annex, facing down towards entrance of building*. **(4)** Side2 - *shot side of entrance, perpendicular to motion, facing east*.

Filming Status: Complete,

Expectations: Will not use footage.

10. Rock Concert (Not Shot Yet)

(Pedestrian Entrance/Exit)

Shot Description: Will shoot the Entrance/Exit of a rock concert. We expect a steady flow into the venue, and *light, medium* and *heavy* outpour. Three Angles: **(1)** Front - *eye level and head on facing the people exiting*, **(2)** Side - *eye level and facing perpendicular to motion*.**(3)** Oblique - *Shoot from high above and wider angle of exit.*

Filming Status: Venue Not found

11. Intersection01

(Vehicular Intersection)

Shot Description: Shot at the intersection of La Jolla Village Drive and Villa La Jolla Drive over the course of several days during mid-summer. The three angles observed were (1) Front - *ground-level center facing west*, (2) Side - *ground-level facing south*, and (3) Oblique - *birds eye view facing east*.



Filming Status:

Still Shots -

- (1) Front Angle - *Complete*
- (2) Side Angle - *Complete*
- (3) Oblique Angle - *Complete*

Panning Shots -

- (1) Front Angle - *Complete*
- (2) Side Angle - *Complete*
- (3) Oblique - *Complete*

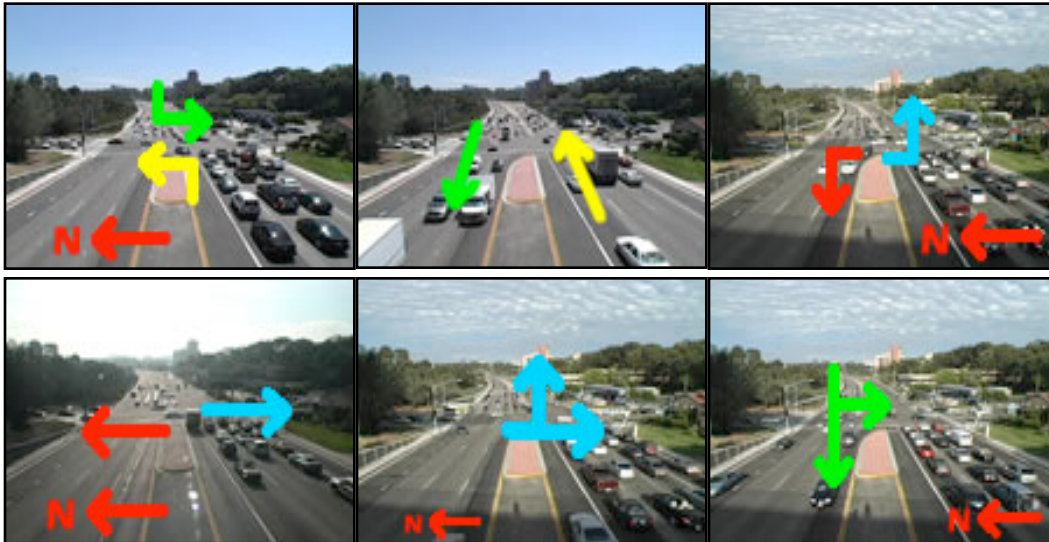
Traffic-Level Classifier Description: Initially we attempted to create a congestion level classifier for Intersection01. This was unsuccessful due to several reasons. (1) there may be many many cars in the interection, however if they are not moving then the intersection congestion is classified as low. (2) no clear definition of 'high', 'medium', and 'low' - lack of language to describe intersection congestion.

Experimental results: ~60% accurate.

Intersection State Classifier Description: Discriminates between six states of the intersection using only the *Oblique* angle. **Eight States:** (1) East-West-Thru (EWT) - *initially east bound and west bound cars are continuing straight through intersection*, (2) East-West-Left-Turn (EWLT) - *initially east and west bound vehicles are turning left through intersection*, (3) North-South-Thru (NST) - *initially north and south bound vehicles continue straight through intersection*, (4) North-South-Left-Turn (NSLT) - *initially north and south bound traffic are turning left through interseciton*, (5) South-Both-Left-Thru (SBLT) - *initially south bound vehicles are turning left and continuing through interseciton* , (6) West-Both-Left-Thru (WBLT) - *initially west-bound vehicles are turning left and continuing straight through intersection*, (7) East-Both-Left-Thru (EBLT) - *initially east-bound vehicles are turning left and continuing straight through*

intersection, and (8) North-Both-Left-Thru (NBLT) - initially north-bound vehicles are turning left and continuing straight through intersection.

Shown below are depictions of the first six states.



(1) Oblique - Initial Experiment Results:

md_error_mean	=	0.1676 = 83% Accurate
kl_error_mean	=	0.1579 = 84%
kl_svmerror_mean	=	0.1809 = 82%
ikl_error_mean	=	0.3234 = 68%
ikl_svmerror_mean	=	0.2713 = 73%

12. Intersection02

(Vehicular Intersection)

Shot Description: Shot at the Intersection of La Jolla Village Square and Nobel Dr. Three Angles: **(1)** Front - *west side of intersection facing east, eye level, set up on island which separates east and west moving vehicles.* **(2)** Side - *north side facing south, eye-level, vehicles exiting plaza on our right, vehicles entering on left side,.* **(3)** Oblique - *Shot from nearby hill on northwest corner facing southeast. Shot a little higher than eye level.*

Filming Status:

Still Shots -

(1) Oblique Angle - *Complete*

(2) Front Angle - *Complete*

(3) Side Angle - *Complete*

Panning Shots -

(1) Oblique Angle - *Complete*

(2) Front Angle - *Complete*

(3) Side Angle - *Complete*

Classifiers: No Classifiers have been built using this footage. This scene may not contain enough variation in traffic flow to test the Traffic-Levels (congestion) classifier.

13. Freeway Entrance

(Vehicular Entrance/Exit)

Shot Description: Shot on the South-West region of the intersection of Interstate 5 Freeway and Genesse. Three Angles: **(1)** Front - *eye level, set up on island which forks freeway entrance lane and street lanes, facing west* **(2)** Side - *eye-level same location as front angle, however facing south,* **(3)** Oblique - *Shot from nearby hill on south side of first two angles, facing northwest.*



Filming Status:

Still Shots -

- (1) Oblique Angle - *Complete*
- (2) Front Angle - *Complete*
- (3) Side Angle - *Complete*

Panning Shots - none

Angle Classifier Description: Discriminating between *front*, *side*, and *oblique* angles. One Experiment to distinguish between the three angles.

Results:

md_error_mean	=	0.0333	=	97%	Accurate.
kl_error_mean	=	0.1430	=	86%	
kl_svmerror_mean	=	0.2294	=	78%	
ikl_error_mean	=	0.0987	=	91%	
ikl_svmerror_mean	=	0.1152	=	89%	

Future: Will crop video for the purpose of a Traffic-Level Classifier.

14. Parking Structure

(Vehicular Entrance/Exit)

Shot Description: Gilman Parking Structure is located at the south-east edge of UCSD. During the school year parking spaces usually fill up by 8 or 9 am. Therefore the largest incoming flux is early in the morning.. Mid day there is a steady in and out of cars through the south facing exit. Around 4-5pm there is peak exiting as this is when many get off work.

Preliminary shots were taken from a few angles. The most promising angles are those from **(1)** the east side of the south exit, **(2)** In the center of the exit, facing the parking structure so that we are facing cars exiting, **(3)** From the second or third story shooting down upon the cars.

Filming Status: Preliminary stage.

Expectations: Although this may be a good scene to create a database on vehicular entrances/exits, there is a new parking structure opening up south adjacent to RIMAC that will be a larger. This may be a more ideal location. Furthermore, Horton Plaza in downtown San Diego has a large parking structure that may be worth examining for potential footage.

Ongoing Research

From here we will continue to film and build classifiers on the video we currently have. We have noticed that a bug in our experimentation: we can achieve higher accuracy if we perform experiments entirely from scratch (i.e. recomposing the SVMs and Nearest Neighbors from scratch) after each change in the experimental parameters. We fixed this bug by coding a deletion of the previous classifier before building a new one in experimenting.

Although there is a desire to clean up the training data to build a more accurate classifier, those outliers that fall in between classes are retained. This

gives us more accurate feedback on the ability to model real world data.

We will also attempt to mix angles. We desire for a classifier to discriminate motion in only one angle. Additionally it is desirable for the classifier to discriminate between angles themselves.

For Example: If congestion can be accurately classified from two angles independently (e.g. high, medium and low from oblique and front angles, giving a total of 3 classes for each of two classifiers.), then these should be able to be combined into one classifier (e.g. 6 classes in one classifier). This should not affect accuracy.

Conclusions

Shown here are initial results from experiments. Accuracy is expected to increase after varying parameters in classifier training. Kullback-Leibler divergence generally is more accurate. However, the Martin distance proved more accurate compared to Kullback-Leibler in Angle '(3)Oblique' of the '01.Sidewalk01' scene (md_error vs. kl_error).

Before experimentation we suspected that our *bird's eye view* would garner better accuracy due to the nature of dynamic texture modeling. This seems to be the case with initial experimental results. However, we need to test this hypothesis more extensively.

Finding very rough optimal values for our experimental parameters using half the data saves time.

Filming on campus during the summer is not suitable. There are not enough people to obtain a large variation in traffic flow.

I'd like to experiment with using these classifiers to determine ground speed of observers (possibly the velocity too).

This technology may be a suitable tool in other fields such as neuroscience (e.g. where dynamic mapping of optical stimulus to optical nerve transmission is an interesting problem for curing blindness).

ⁱ Antoni B. Chan and Nuno Vasconcelos. Classification and Retrieval of Traffic Video Using Auto-Regressive Stochastic Processes. SVCL-UCSD, 2005.

ⁱⁱ Ibid.

ⁱⁱⁱ http://www.calit2.net/newsroom/print_page.php?id=928