IDEA

Innovations Deserving
Exploratory Analysis Programs

Transit IDEA Program

Enhancing Safety and Security of Transit Systems using Computer Vision

Final Report for
Transit IDEA Project 80

Prepared by:
Dimitris N. Metaxas
Rutgers University

September 2017

TRB TRANSPORTATION RESEARCH BOARD
The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Innovations Deserving Exploratory Analysis (IDEA) Programs Managed by the Transportation Research Board

This IDEA project was funded by the Transit IDEA Program.

The TRB currently manages the following three IDEA programs:

- The NCHRP IDEA Program, which focuses on advances in the design, construction, and maintenance of highway systems, is funded by American Association of State Highway and Transportation Officials (AASHTO) as part of the National Cooperative Highway Research Program (NCHRP).
- The Rail Safety IDEA Program currently focuses on innovative approaches for improving railroad safety or performance. The program is currently funded by the Federal Railroad Administration (FRA). The program was previously jointly funded by the Federal Motor Carrier Safety Administration (FMCSA) and the FRA.
- The Transit IDEA Program, which supports development and testing of innovative concepts and methods for advancing transit practice, is funded by the Federal Transit Administration (FTA) as part of the Transit Cooperative Research Program (TCRP).

Management of the three IDEA programs is coordinated to promote the development and testing of innovative concepts, methods, and technologies.

For information on the IDEA programs, check the IDEA website (www.trb.org/idea). For questions, contact the IDEA programs office by telephone at (202) 334-3310.

IDEA Programs
Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001

Enhancing Safety and Security of Transit Systems using Computer Vision

IDEA Program Final Report

Transit IDEA-80 Project

Prepared for the IDEA Program
Transportation Research Board
The National Academies

Dimitris N. Metaxas
Rutgers University
Submittal Date
9/11/2017

TABLE OF CONTENT

EXECUTIVE SUMMARY

Detecting distraction and fatigue in truck drivers is very important for preventing accidents and improving their performance. To this day there is no available fatigue monitoring system that is reliable, easy to use, non-intrusive and inexpensive. We have recently developed a novel computer vision based technology that is based on a single camera and uses deformable model technology to track from video the geometry of any person's face, his/her expressions and measure slow eyelid closure (PERCLOS) which is directly related to fatigue. Our system is non-intrusive and can work with both visible and infrared light regardless of the driver background which changes over time during driving. What makes our system reliable is that its tracks the whole face and therefore can reliably detect the eyes and the eyelids. Our system is also potentially cost-effective since it is based on an off-the-shelf computer and a single camera.

Our approach to analyze drivers' faces involves three steps. In Step 1, the orientation of the subject's head needs to be estimated ("global" head pose). In Step 2, the position and shape of the subject's eyes need to be estimated, and in Step 3, the eyes' openings and closings need to be captured/tracked. Most techniques in the literature dealing with the eye localization and tracking (e.g., eye-gaze tracking), assume that the pose of the head is fixed and frontal. Although this might be a valid assumption in certain applications (e.g., medical applications), it is a rather restrictive and invalid assumption in many other applications and under natural conditions, e.g. in a car and while the subject is driving. To build a robust PERCLOS detection system, all of the above problems need to be tackled.

Our proposed system is based on a single CCD camera connected to a computer. The system will be able to detect and analyze in real time the face and its movements, under varying head poses and lighting conditions, and to detect eye blinking and PERCLOS of the person in front of the camera. The rigid movement and the 3D pose of the head as well as the deformable motion of the face, and specifically the eyes, will be handled by a combination of 2D and 3D tracking approaches, already developed by Dr. Metaxas group. Our system works with both visible and infrared light regardless of the driver's scene background, which changes over time during driving. We test our system for detecting fatigue based on data that have been already collected and will be provided by our longtime collaborator and expert on fatigue, Dr Dinges from the University of Pennsylvania.

Our face and eye tracking framework combines Active Shape Models with a predictive face aspect (discriminative) model to track features across large head movements, and runs in real time. An Active Shape Model (ASM) is a deformable model for shape registration that detects facial features by combining prior shape information with the observed image data. Our framework represents various head poses by multiple 2D shape models and accounts for large head rotations by dynamically switching between them. Our switching variable (the current model to use) is discriminatively predicted from the SIFT descriptors computed over the bounding box of the low-resolution face image.

We use the results of our 2D tracking framework to deform a 3D face mesh in order to guide the feature tracking across partial face occlusion, full profile head movement and self-occlusion of eye features. The 2D – 3D coupling ensures smooth and robust tracking of eye features across varied illumination and appearance changes due to glasses. After estimating the exact 3D head pose and the eyes position of the subject, we propose to apply region-based techniques inside the eyes' boundaries that indicate closings and openings. More specifically, we will estimate the (locally) normalized histograms of oriented gradients to discriminate between the states "eyes closed" from "eyes open". For infrared image, we train a Support Vector Machine (SVM) model to classify the eyes' status into two categories, namely "open" or "closed". The detection of these states will enable us to determine useful features, such as the frequency, the velocity, and the duration of eye blinking.

Our system has the potential to revolutionize fatigue detection and result in safer driving by automatically alerting the driver. It can also be used in other very important applications such as drunk driver detection since it tracks the whole face and it is easy to spot abnormal facial behavior.

This project provides automobile drivers the ability to detect distractions and fatigue during driving in order to prevent accidents. The system is based on a single camera aimed at the driver's face and will provide a warning signal when the driver is distracted or fatigued.   The Rutgers team develops the system and test the method and advance it to a prototype for application and potential commercialization. The testing takes place on buses at Southern Pennsylvania Transportation Authority (SEPTA) and bus simulator from Metropolitan Transportation Authority of NYC (MTA).

The three primary objectives of this research project are:

1.) Develop a prototype of driver facial monitoring system via porting the software onto a portable processor to create a hardware system.

2.) Initial prototype testing using existing or acquired videos to test the accuracy of the tracking ability of the prototype with a variety of facial features, including glasses, different skin tones, facial hair; with different levels of illumination; and with various degrees of facial movement.

3.) Public beta testing of prototype on buses at SEPTA and MTA to evaluate the accuracy in terms of: (a.) validity – that the system accurately measures what it is supposed to measure, and (b.) sensitivity – that the system picks up distraction and fatigue with minimal false positives and negatives.

THE PROGRESS IN PHASE 1

In Phase 1, the goal is to design the face analysis algorithm to estimate driver distractions and set up a portable system to collect the data from the real cases and do a first evaluation of its performance. Phase 1 consists of the following four tasks which we have completed with no deviations from the contract:

1) Porting of Software onto a portable Processor

2) Evaluation of Robustness of the Tracking Algorithm

3) Evaluation of System Performance with Driver Head Rotation

4) Evaluation of System for Distraction and Fatigue

Based on discussions with SEPTA in Philadelphia (Mr James Stevens), we were able to install our device on a SEPTA bus. We collected data and conducted the evaluation as stated above. The details of the portable face analysis system and the evaluation are presented in section 1.

Based on our analysis results from the SEPTA bus, we have found that the system performs very well if the data quality is good, i.e., if the face of the driver is visible. However, there were some cases which depended on the placement of the camera and the selection of the type of camera that impacted on the system performance. We propose and describe in Section 2, to make small modifications to the camera location and the type of camera to complete the evaluation of our system in Stage 2.

In the following, we provide details on our Phase 1 accomplishments.

ACCOMPLISHMENTS DURING PHASE 1

In the following we summarize the accomplishments for each Quarter based on the corresponding TASKS, followed by details on each of the TASKS.

Accomplishments during the First Quarter
During this quarter, we completed Task 1 as described in the Contract. We successfully port our software onto a portable processor. The processor uses an HD 60fps video IR camera as input and a digital screen for output. In addition, the processor outputs the video and the analysis to its hard drive for evaluation purposes. The processor, camera, screen and hard drive are powered based on a bus's 12V outlet and are packaged into a portable protective case to fit in a bus's storage compartment located behind the driver.

We prepared the Beta testing of our algorithm by porting it to a portable computer. The main accomplishment was the development of a portable device with our driving monitoring system ported. This accomplishment allowed us to put the system on buses to test the robustness of our approach in real driving scenarios and collect more data which can be used to boost the current performance of our software.

Two subtasks were necessary to accomplish. The first was to develop a portable working destruction driving monitoring system. We did research to determine which portable computer is appropriate to which we will hook a camera and a small monitor, while at the same time can be powered by the bus power system. The additional devices we looked included a portable monitor and a camera, in terms of the cost, device hardware specification, size.

Once we developed the device, the second subtask was to port the current driver monitoring system on an off the shelf processor, including compiling the necessary library and re-writing our face tracking codes so as to be compatible with the portable processor. In addition, we also tested the face tracking system on the portable device and reduced the computational complexity to be able to run the code. The modified program requires ~100MB of storage on the hard drive and ~120MB of main memory during run time. These allow the running of our system at 30Hz,i.e., both processing and outputting of the results on the screen.

During this period, we were in close contact with Mr. Jim Stevens (SEPTA) in order to get permission to put our device on a SEPTA bus. The details for each of the subtasks are described below.

Subtask One: The Initial Facial Tracking Algorithm

Our face tracking software is based on Computer Vision technologies and combines different model-based 2D and 3D tracking approaches. We defined 66 key points (landmarks) to locate the facial parts, such as eyes, eyebrow, nose, and mouth, as shown in figure 1. Landmark localization addresses the problem of matching a group of predefined 2D landmarks to a given facial image. Landmark tracking is to continuously capture the landmarks in an image sequence. Such tasks are prerequisite for analysis of face activity that requires accurate landmark positions, such as face recognition, facial expression analysis, facial modeling etc.

For the face and facial features' localization and tracking, we use Active Shape Models [1] with a predictive face aspect (discriminative) model that tracks features across large head movements, and runs close to real time. Active Shape Model (ASM) is a deformable model for shape registration that detects facial features by combining prior shape information with the observed image data. Our framework represents various head poses by multiple 2D shape models and accounts for large head rotations by dynamically switching between them. Our switching variable (the current model to use) is discriminatively predicted from the SIFT descriptors computed over the bounding box of the low resolution face image.

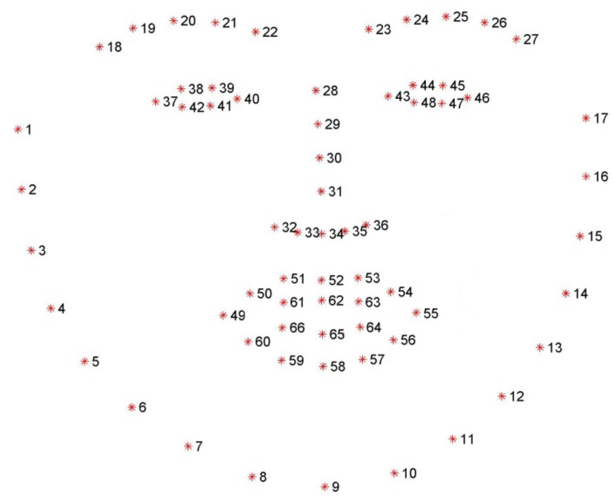

Figure 1. Pre-defined 66 Facial keypoints

We then use the above 2D tracking framework to deform a 3D face mesh in order to guide the feature tracking across partial face occlusion, full profile head movement and self-occlusion of eye features. The 2D-3D coupling ensures smooth and robust tracking of eye features across varied illumination and appearance changes due to glasses.

After estimating the exact 3D head pose and the eyes position of the subject, we applied region-based techniques inside the eyes' boundaries that indicate closings and openings. More specifically, we will estimate the (locally) normalized histograms of oriented gradients to discriminate between the two different states of "eyes closed" and "eyes open". The detection of these states enables us to determine useful features, such as the frequency, the velocity, and the duration of eye blinking. This way we are able to measure robustly PERCLOS and fatigue (slow eyelid closure) and distractions based on head orientation.

SubTask 2: The Portable Driver Monitoring System

Figure 2 illustrates the portable device we developed to monitor a bus driver. The processor is the Intel NUC Kit at 4.53 x 4.37 x 1.92 inches. We installed a windows 64-bit OS on the processor to simplified the work of the code transferring.

The processor outputs the video and the analysis results to its hard drive for evaluation purposes. We set up two SSD (up to 500GB) drives for storing the analyzed result and collecting data which can be added to train our model.

The processor allows the use of two types of cameras we considered to use as input in our monitoring program. The first is a visible light High definition 30fps typical camera and a 30fps High Definition IR video camera. Our current blinking detection model is trained on RGB data set. So, based on discussion with SEPTA we plan to try visible light camera in the initial stage. After we collect more data, we will upgrade the model to the IR camera version which deals better with light variations than a visible light camera, especially during the night in low lighting condition.

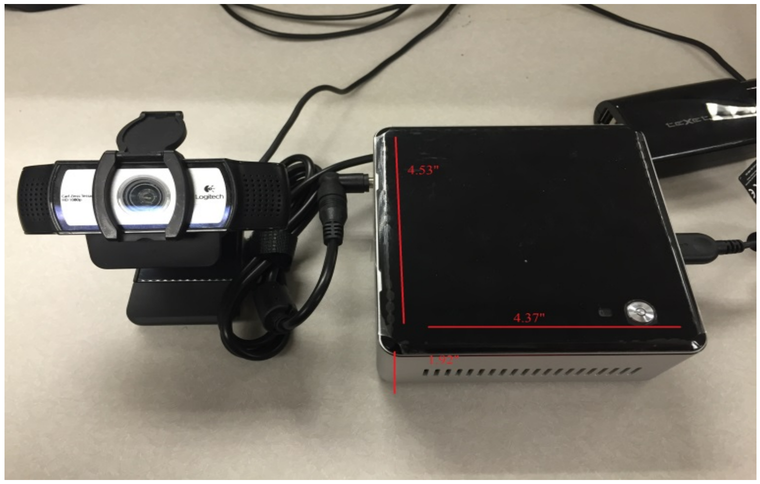

Figure 2: Depiction of our portable driving monitoring system parts.

The whole system is powered by a transformer cable that outputs 19V using as input the 12V bus power outlet. The computer will be packaged into a portable protective case to fit in a bus's compartment behind or above the driver seat. This case was given to be used by SEPTA.

In addition, the processor is equipped with WIFI 802.11ac and Bluetooth so the monitoring system has the capability to wirelessly transmit the data when a fatigue or distraction event happens to the depot for monitoring by the driver's supervisor. When the bus goes to the depot at the end of a day's service all the video and its analysis will be uploaded to a server which will be available to us for further statistical analysis. This is our monitoring system.

Accomplishments during the Second Quarter

During this quarter, we conducted two kinds of evaluations, which consist of the evaluation of the robustness of the tracking algorithm and the Evaluation of the System Performance with Driver Head Rotation. The data collected in the experimental setting helped us to conduct solid evaluation and modify our algorithm so that the facial tracking system can work on a bus during its normal operating conditions.

Specifically, we first made some software modifications to our face tracking system in order to be able to work on a bus. These changes amounted to being able to deal with varying lighting, extreme head pose variations up to 70 degrees. With the help of SEPTA, we then set up our driver face monitoring system on a SEPTA bus to collect driver data and also stored the analyzed results for preliminary evaluation. We used these data collected on SEPTA buses to find the limitations, modify the algorithm and refine the experimental settings.

Evaluation of the System Robustness

We tested our 2D facial feature localization and tracking, our 3D face mask extraction and tracking, and our eyelids tracking methods on a variety of collected videos. For our 2D facial feature extraction method, we tested our system under different lighting conditions, low resolution and in several faces, including extreme cases of facial expressions, head rotations, facial hair and eyeglasses.

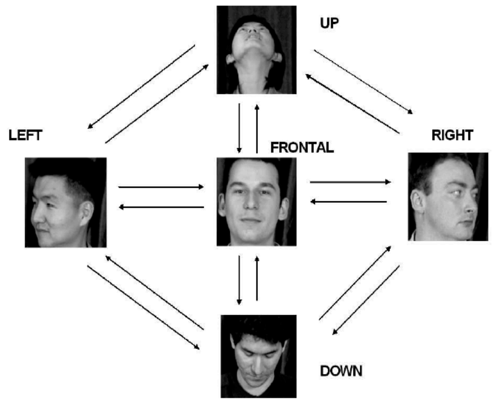

Figure3. Extreme head poses our system can handle

Since the tracking accuracy depends also on the frame-rate of the camera, we improved our tracking system to be able to work at higher than 30fps video sequences. Also, we incorporated more ASM models in our system, to handle extreme head poses; so far, our findings suggest that the 5 pose ASM models (frontal, left, right, upward and downward head pose) are sufficient to handle also extreme poses.

We then tested our 3D face pose tracker, using the extracted 2D tracking results, and we made all modifications needed to increase the tracker's efficiency. Since the 3D face mask fitting on the image depends on the extracted 2D features, we were able to obtain accurate 3D features in all the cases that our 2D facial feature extraction and tracking method can handle. Similarly to our 2D face tracking method, we tested our 3D method in extreme head poses as shown in Figure 3.

Evaluation of System Performance with Driver Head Rotation

We tested the overall system performance which will be computing all the necessary facial measurements at rates faster than 30fps so that it can be operational on a SEPTA bus. We especially tested the system performance when the driver makes moves that make the face disappear from the viewpoint, the driver head rotates more than 70 degrees with respect to the camera or makes expressions. These facial computations are based on features extracted the eyebrows, the eyes, the mouth and the 3D head rotation and translation. The accuracy has already been tested in the lab based our VICON system and the facial tracking accuracy is less then 4pixels at 1080p resolution. So we are confident, the same performance will be continued with the rest of the data we will collect on the SEPTA bus. Initially in the lab when we tested our results we measured 1mm accuracy since we also used a VICON system. However, in real data, where there is no possibility of using a range sensor, we measure deviation between a human marked point and the corresponding model point. For example, landmark testing points include the corners of the eyes and the corners of the mouth.

Setting the Device on a SEPTA Bus

In the second quarter, we used a camera location approved by the SEPTA bus driver Union, on the SEPTA bus. The camera is connected to our portable processor through a USB port. The video frame are recorded by this camera in RGB format (non-infrared) at 30 fps, and our face tracking system is
compatible with such data. Figure 4 shows the camera position using a red square.

Figure 4. Camera on the bus (marked by red square)

As shown in Figure 5, the data collected using this location of the camera introduce challenging scenarios, such as occlusion, side facial views, extremely varying background and often bad lighting. In the third quarter, we collected more data from the SEPTA bus and fixed possible software limitations of the face tracking part of the software.

Figure 5. Sample Data Collected on Septa Bus

Accomplishments during the Third Quarter

During this quarter, with the data collected on the SEPTA buses, we conducted Evaluation of the driver monitoring System for Distraction and Fatigue. Initial testing of the Rutgers software to detect driver distraction and PERCLOS/Yawning/Head movement on a variety of existing videos as well as pilot data we will acquire from a simulator and training bus at SEPTA. The PERCLOS/Yawning/Head movement detection over a period of a minute will be used as a possible sign of fatigue. The initial SEPTA data will be done based on the application and approval of the appropriate IRB. We plan to collect video and its analysis from at least 8 different drivers who drive a single bus over the course of a month.

First, we had to modify our face tracker to be able to analyze the challenges introduced by the data we collected on the SEPTA bus. In the following, we describe the details of the modified algorithm and the testing conducted on the collected data from SEPTA bus. We modified our algorithm for driver distraction based on data collected on the SEPTA bus. We applied our analysis system to track the drivers’ faces and analyze their behaviors.

Improvement for Facial Keypoints Detection
In third quarter, we made changes to the initial algorithm and developed the real-time face tracking and analysis system which can simultaneously handle the following processes: face detection, pose-free landmark localization, face tracking, detection of driver distraction and PERCLOS/Yawning/Head movement. In the face detection stage, a group sparse learning approach [2] is employed to automatically select the most salient facial landmarks. The group sparse learning based method is proposed to achieve real-time performance for tracking, which dramatically decreases the number of potential landmark areas and still preserves the detection effectiveness. In the facial landmarks localization stage, a two-phase cascaded deformable shape model is adopted to localize landmarks with large head pose variations. The first phase a global optimum is rapidly achieved through mean-shift local search with a constrained local model. In some cases, merely mean-shift strategy cannot find the correct facial landmark positions because the global constrain in the first phase does not guarantee fitting each facial component exactly. In the second phase, the refinement to better estimate subtle shape variations is achieved by the use of a component-wise active contours approach. The pose-free facial landmarks are initialized by a 3D face shape model and the Procrustes analysis [3] [4] where a pictorial structure [5] is introduced to organize the landmarks. Moreover, we apply the algorithm [6] to deal with partial face occlusion to enhance the system robustness.

Analysis of the SEPTA Bus Data
We produced a portable version of the tracking system on Windows operating system, including ".exe" file (<5 MB) with necessary modules (< 100MB). We collaborated with the Southeastern Pennsylvania Transportation Authority (SEPTA) and put our devices on the buses. The portable device is made up of a small computer and a single web camera. Figure 1 depicts our devices for face tracking on the bus. The size of the computer is about 4.53"x4.37''x1.92" and the web camera is a Logitech 1080p with recording speed of 30 fps.

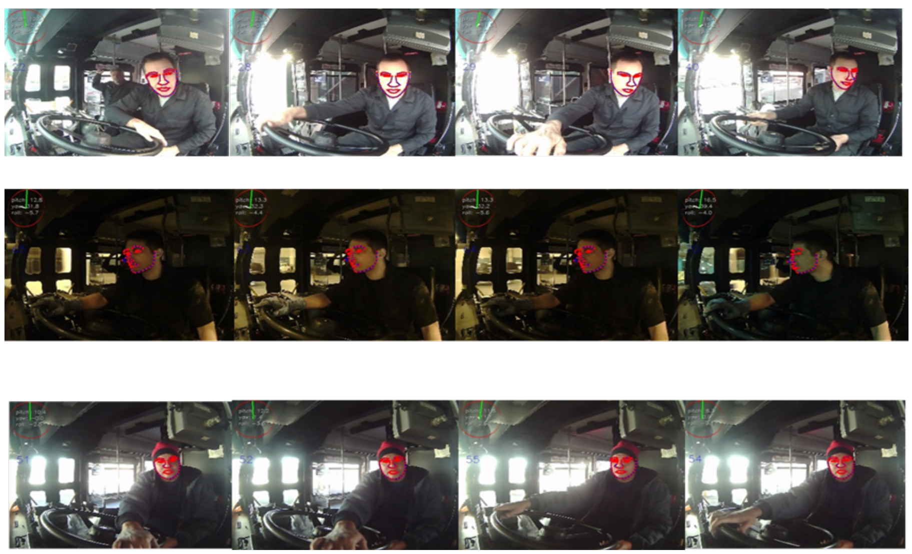

Figure 6. Face Tracking Samples of Drivers Face on the Buses

Figure 6 shows analysis samples for the drivers' faces on the SEPTA bus. The masks are plotted on the faces based on the detected landmarks. Based on the landmarks, the head pose can be estimated. Based on the head pose estimation, we defined **distraction** as follows:
During driving, if the driver keeps their head turned to the side (more than 20 degrees) for a long time over 4secs), it can be a potential dangerous situation and our system will remind the drivers to turn back and look straight.

We also analyzed the eyes area. If a driver has slow eye lid closure frequently, it can be a sign of fatigue and our system reports it (10ms instead of the typical 3ms). However, we did not test fatigue in Phase 1.

The total length of the recorded videos is ~120 hours. One third of those videos are analyzable and we call them the positive samples. As shown in Figure 7, the **positive** samples give us clear faces (front or near-front faces, which happens when drivers are looking the left mirror or turn a little bit to the left. The success rate in these cases is over 97%. We define success as follows: If the mask correctly estimates the direction of the head and correctly fits to the face as seen visually. As shown in Figure 8, three reasons lead a video sequence to be a **negative** sample: 1) no driver occurs in the video; 2) faces cannot be seen clearly due to the dark light; 3) videos just contain side faces due to the improper camera position.

Distraction recognition is equally successful if the tracking is successful.

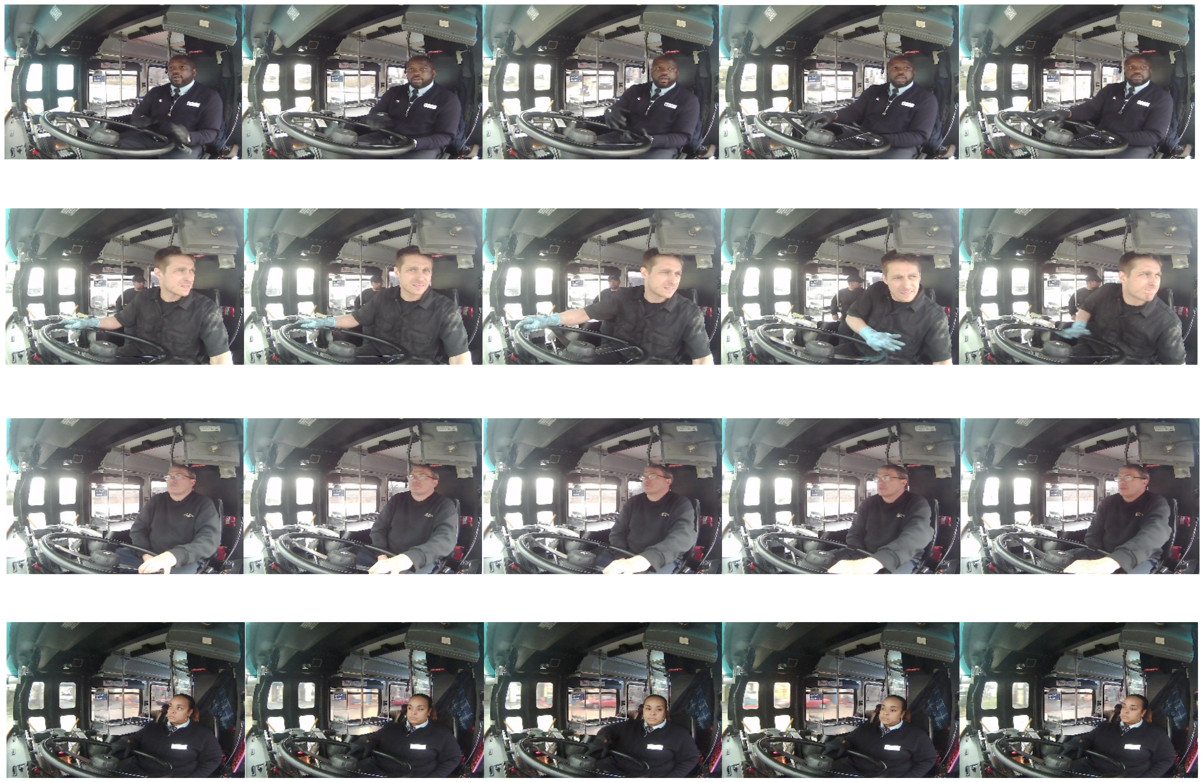

Figure 7. The samples of positive videos in which our system can track the faces, analyze the head pose, expression and detect blinking (eye lid closure and opening).

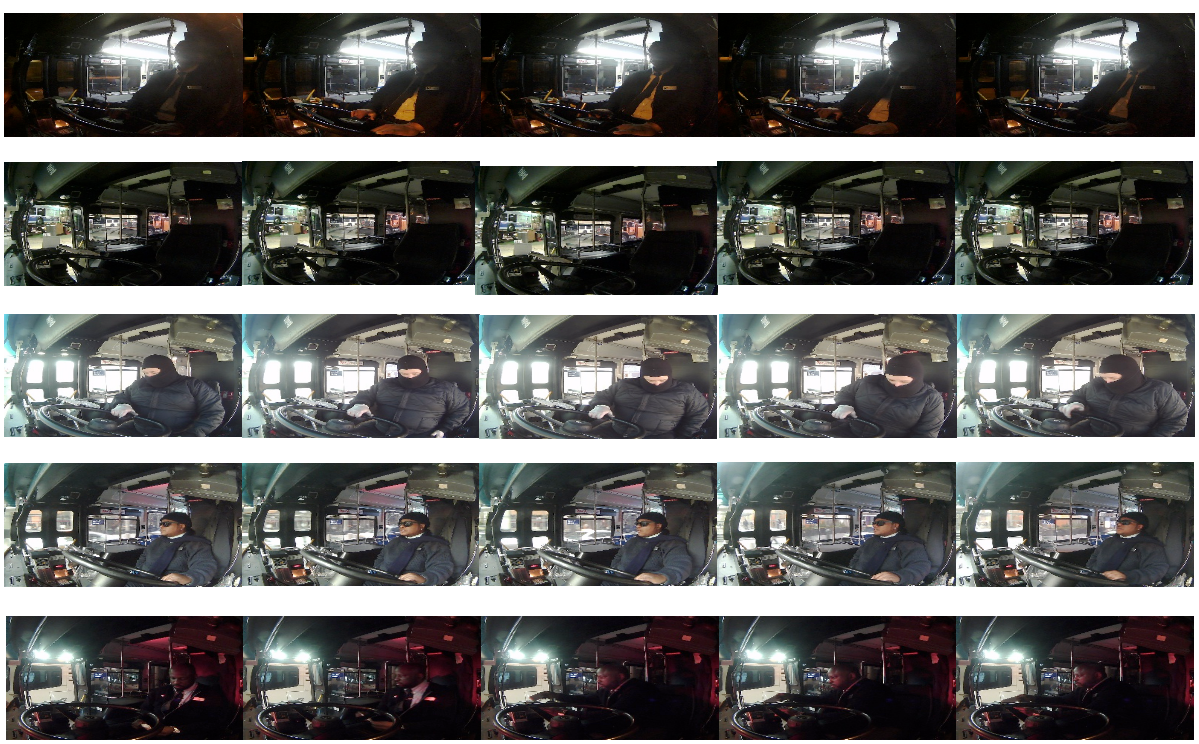

Figure 8. Negative Videos

Our algorithm does recognize the negative cases since it is a model-based approach. In other words, we know when the model does not fit to the driver's face and fits elsewhere in the image or not due to lighting or other conditions.

While Stage 1 was successful and we did not deviate from the proposed TASKS, we need to make the following proposed hardware and software changes in order to make our system more robust and able to work 24/7.

Problems addressed by the data from the real cases

According to the negative samples described in previous section, we have collected various data on the SEPTA bus which includes several cases that our current tracking system cannot handle well. These are caused by the type and placement of the camera and they include: a) dark images when there is low light where there is no visible face, b) face views where face is not visible and is heavily occluded when for example it turns to the right.

To deal with the hard cases, we propose to collect new data, perform final modifications and testing of the system during day and night conditions based on
1) a modified camera location right below the center of the drivers' visor,
2) a new IR camera, and
3) modification to the software and evaluate the system on new data

We now give a description of the proposed modifications and testing of the system:

1. ***Change the camera position and use an infrared camera for day and night driving***

Figure 4. shows the camera in the left corner of the bus, which gives us clear images of the left side of the face due to its position.

In order to improve our face data, we need to change the camera position to obtain the front or near front face of the driver when he/she is looking ahead. Figure 9 shows potential positions of the camera, where those positions are marked by a green square. Based on discussions with Mr. James Stevens from SEPTA, the location of the middle of the visor will be selected as it provides unobstructed views of the driver's face.

Figure 9. Potential camera positions

2. ***Collecting more driving images on the bus and train the face detection model on the SEPTA data.***

Due to the current camera placement, we often have degenerate views (when the driver looks to the right) or dark images at night which result in the face tracking system to fail. In our current data, the background (non-faces) area shows mostly fixed patterns, such as the door, windows or the seats on the bus, which are distinguishable compared to a driver's face.

Based on the new camera location we will improve our training sample for the background (negative samples) and the face (improved left, center and right views) which promises to improve the performance of our current tracking system. We will also add new features by estimating the pupil location, to improve the distraction and fatigue detection.

We are in close contact with SEPTA (Mr. James Stevens) to change the location of the camera and convert it to an infrared camera so we can have a better location and 24/7 face tracking ability to see the driver's face. We will collect once the infrared camera is installed more data and analyze them and then we expect over 97% correct performance in terms of tracking, distraction and fatigue and no complications due to lighting.

THE PROGRESS IN PHASE2

During this quarter, we modified the current face tracking system in order to improve the robustness. The modifications also allowed us to improve the analysis of the driver's behavior. Our results are based on the tasks described in the contract, including improving the tracking accuracy and robustness of the system, final testing and adaptation of user inter face. This evaluation is first performed in the laboratory and then we deployed the system on the general mass transit population Septa and MTA buses. We produced the statistics on the performance of the system, including false positives and false negatives. Based on the initial testing results we made final software modifications. We then continued to test the system on the general mass transit population buses at SEPTA along with extended hours. During this phase, we also optimized the user interface to the driver. In our approach the system outputs two messages on the screen. The first mentions: Distracted Driving. The second mentions Drowsy Driving or display a cup of coffee. If none of these conditions occur, the system only shows the tracking of the driver's face. Even though we are not creating a product as part of this contract, we will discuss with the drivers in the future, the SEPTA management, relevant automotive companies and will do some literature research on best ways to alert a driver visually. Our team also produced the final driver distraction and PERCLOS/Fatigue statistics based on our software system.

ACCOMPLISHMENTS DURING PHASE 2

According to our results from the SEPTA bus data in Phase 1, there are two main problems in the current face tracking system.

1. The camera viewpoint is to the left of the driver and we have many left-side video due to the improper camera position, which degrades the tracking performance. This way we can't see a face turning to the right. For a robust tracking, it requires a frontal or near frontal face in the initial frame so that the face area can be detected accurately.

2. There are a lot of dark images due to the lack of lighting.

During this quarter, we have addressed the two problems based on data from a different set of busses from the Metropolitan Transportation Authority of NYC (MTA), where the camera position was changed to the middle of the bus. We also modified the system to have an infrared camera to get IR images to ensure operation in low light conditions. We would like to test the system also on a SEPTA bus but was not possible to get the right permissions during Phase2. Therefore, we will be analyzing new data from the MTA.

Adjustment of the Camera Position

As it was not possible to place our camera in the modified position on a SEPTA bus, we communicated with MTA and we paced a camera properly in their simulator. Figure 10 shows the new data from the MTA simulator bus and the tracking results. Compared to the previous data shown in Figure 11, we can capture better frontal faces, which results in improved facial analytics.

Figure 10. Sample data collected from the MTA simulator bus.

Figure 11. The previous data we collected on the Septa Bus. The camera is placed on the left corner of the bus.

Usage of Infrared Camera to Avoid Dark Scenes

In order to overcome the lighting problem, the POINTGREY infrared camera (IR) is used to capture faces in low light situations. The code of our program was modified so that the new camera can be compatible with our system. Figure 12 shows the samples of the faces recorded with the infrared camera where it gives us improved images in low light situations.

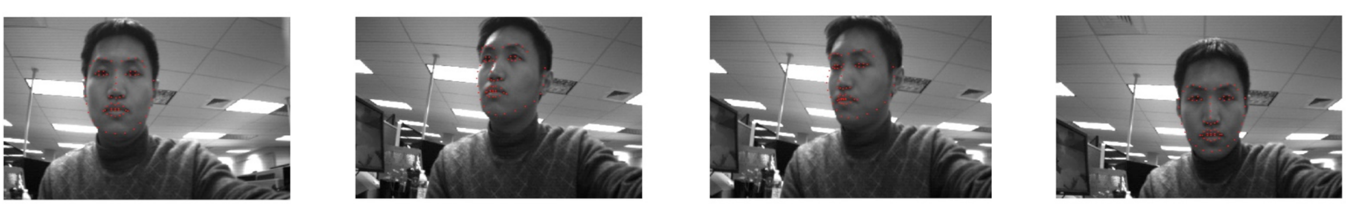

Figure 12. The sample faces captured by infrared camera and the tracking results shown in red.

Head Pose for Distraction Analysis

To measure distraction, we first track the head pose over time, and we use a unified rule for the head pose measured based on three angles: pitch, roll and yaw, as shown in Figure 13. In addition, the time period we selected is 4 seconds and is a variable that can be adjusted.

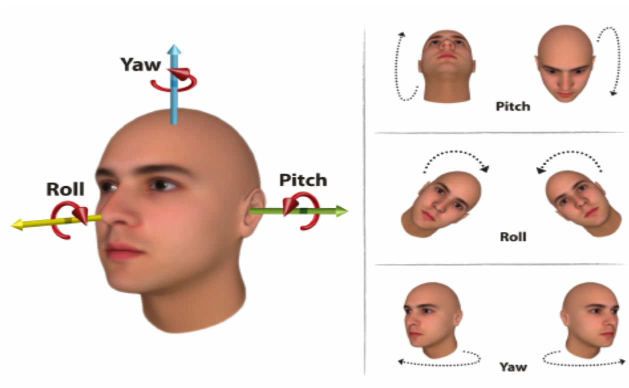

Figure 13. Head Pose Measurement

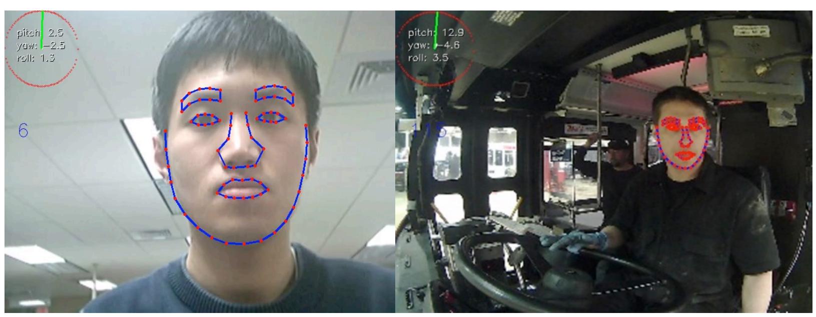

Figure 14. Analyzed frame with head pose information on the top-left corner

Figure 14 shows samples of the analyzed frames. The angles for the head pose are printed on the left top corner in each sub-screen; numbers (blue) indicates the frame ID, and 66 facial key points are plotted. Based on the landmarks, the head poses can be estimated. During driving, if the driver keeps their head turned to the side for a long time, it is a potential dangerous case and our system will remind the driver to turn back and look straight, as shown in Figure 15.

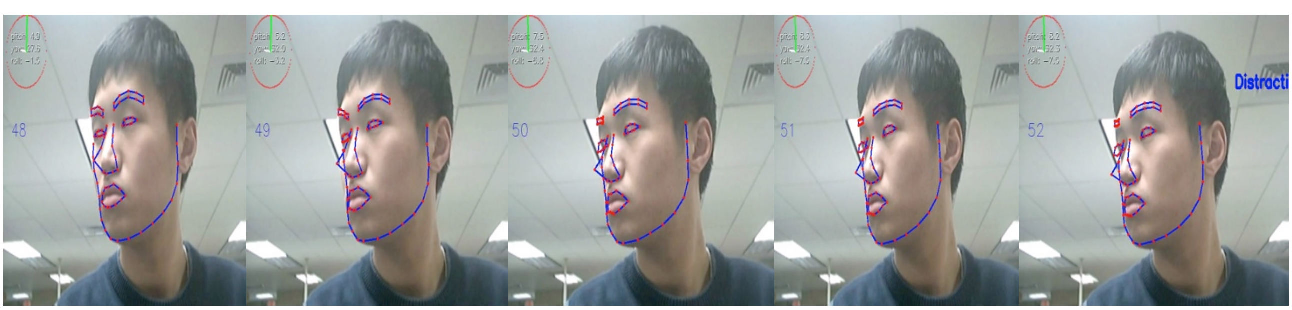

Figure 15. The person keeps looking to the right side for 4 seconds and the system shows the warning “Distraction”.

Fatigue Detection

For fatigue detection, we analyze the eyes area. If the driver shows slow eye blinking frequently, it can be a sign of fatigue. By slow we mean 10ms for closing the eye instead of the typical 3ms. In the future, more robust fatigue analysis will be added, like analyzing the pupil where camera with higher resolution is needed to capture pupil image.

MTA DATA ANALYSIS

We made the modifications on our face analysis systems and collected more data from MTA simulator bus for additional analysis. In this section, we present the quantitative results for distraction. The qualitative samples are also shown for visualization. There were no fatigue cases in the data to analyze.

Data from MTA Simulator

Figure 16 shows the video samples collected by the single camera which is located at the place where the face yaw is in ~0 degrees and the driver is looking straight.

Figure 16. A sample image sequence of a driver looking straight

Data Preprocessing

Generally, the dataset consists of the raw videos during 26 days (6 hours per day approximately). We conducted an automated preprocessing of the raw data to remove frames where less than 40% of the facial model points can be fit to the data. The first step of preprocessing is running the automatic face detection on each frame to check the occurrence of a face in the video. The algorithm works by knowing an average face size and expecting its location to be in the center of the image. If these conditions are not met as shown in Figure 17, the system marks them as garbage frames and does not use them. For these data, we also manually checked these data and our system was able to be 100% successful in recognizing them. Our system defines as garbage frames, those where is no visible more than 40% of the facial points. A sample of garbage videos is shown in figure 17.

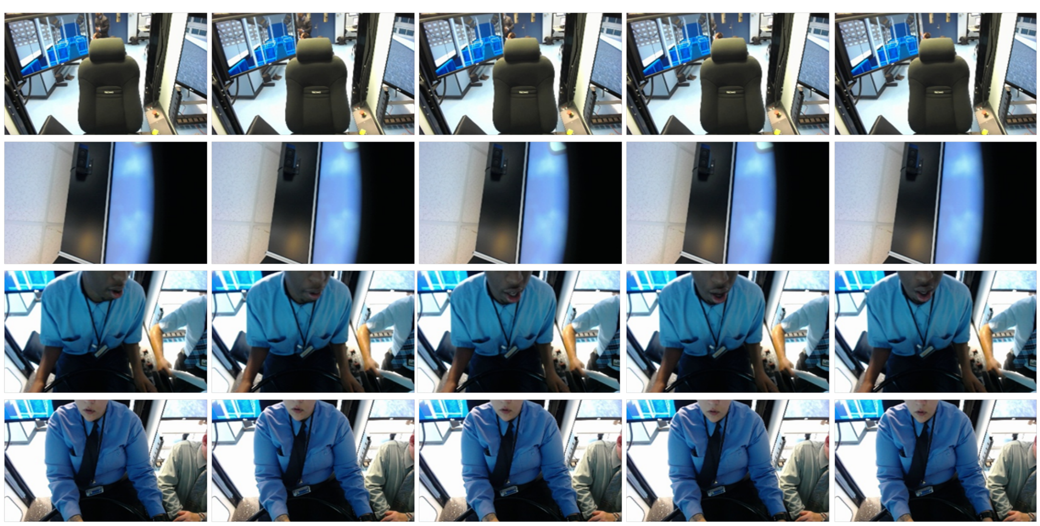

Figure 17 Examples of garbage video sequence caused by no drivers shown and improper camera place

After the preprocessing step, the video clips are suitable for the facial analysis. We captured data over 12 days, and there are

402,138 image frames in total. For validation purposes, we subsample the data using a uniform sampling strategy, resulting in 134,036 image frames.

Driver Distraction Analytics

Since the data are based on a simulator bus (same is the case for the SEPTA bus data), the drivers do not show any signs of fatigue. A worthwhile observation for these simulator bus data is that drivers are asked to make large left and right head turns. When a driver is making a left/right turn the data show that the turn last no more than 4 seconds. If more than 4 seconds, our system will report "distraction". The duration to report distraction is a parameter in our system and can be set for real driving applications and can be set if coupled with GPS information and we plan to investigate in future work.

Based on this set parameter, our system predicts distraction or not (1/0: 0 no distraction and 1 distraction) for each frame with a time stamp of the frame. To test the accuracy of our distraction detection, we randomly selected video segments from 21different drivers and checked the distraction prediction correctness by manual verifying it based on the corresponding video frames. We test distraction or not based on segments of 500 frames and we look for distraction based on continuous turning of the head by more than 120 frames given that our video capture is based on 30fps. This is a second parameter in our system that needs to be set. Also the head needs to be turned more than 30degrees to be considered as turned (this is another third parameter to be set in our system). The following table depicts some of our analysis results. The first column represents a video sequence from many drivers takes during a given period. The second column represents the selected frames from a given driver. The third column depicts the distraction estimation results. The fourth column depicts the distraction ground truth.

Video ID (includes many drivers)	Frames ID (from a single driver)	Distraction Detection Results	Distraction Ground Truth
70038	14097-14544	Have Distraction	1
60848	42987-43434	Have Distraction	1
60848	134640-134789	No Distraction	0
92422	132915-133272	No Distraction	0
92422	66372-66759	Have Distraction	1
64857	19950-20337	No Distraction	0
64857	14415-14772	Have Distraction	1
71154	100875-101232	Have Distraction	1
71154	62349- 62706	No Distraction	0
63336	15132-15489	Have Distraction	1
63336	20724-21081	No Distraction	0
63633	26292-26649	No Distraction	0
63633	27747-20104	Have Distraction	1
62010	117402-117759	No Distraction	0
62010	123264- 123621	Have Distraction	1
62010	107424-107781	Have Distraction	1
62010	108825- 109182	Have Distraction	1
121328	1611- 1968	Have Distraction	1
84131	99603-99960	Have Distraction	1
63633	163923-164280	Have Distraction	1
60848	**229425-230076**	**No Distraction**	**1**

According to the table, we can safely draw the conclusion that if the head pose estimation is reliable and accurate (our results show 97% accuracy from Phase 1), the distraction detection is 100% accurate. In order to set the distraction parameter accurate, we need in future work to integrate our system with a portable USB based GPS device which will allow us to integrate information such as speed, and vehicle turning information.

The last row of the table shows the only case we found that the system failed to report distraction correctly. This was caused by failure in the face tracking as shown in Figure 18. In this video, the lack of appearance of the driver's face resulted in the system tracking the face of the other person in the video. Due to the presence of the second persons face, this video was not marked as garbage. A case like this can be easily modified by changing the field of view of the camera to be narrower. No need to make changes to the software, since it already rejects a face whose aspect ratio is different from the one already tracked.

Figure 18. The single failure case to detect distraction due to facial tracking error.

CONCLUSIONS

Our system has the potential to revolutionize fatigue and driving distraction, resulting in a safer driving by automatically alerting the driver. It can also be used in other very important applications such as drunk driver detection, and driver recognition since it tracks the whole face and it is easy to spot abnormal facial behavior and identity.

1. A major advantage of our system, as mentioned above, includes the accurate driver's eyelid tracking under different lighting conditions, head poses and backgrounds. Our system was tested in real conditions, including driving at night, and can be used with both visible and infrared CCD cameras.

2. Another major advantage of our system is that it is only based on an off the shelf single camera and computer and therefore it is very easy to deploy and potentially very cheap, under $500, if mass-produced.

3. The system has key parameters that can be set to detect accurately distraction and fatigue based on a robust facial tracker.

4. The system knows automatically when the tracking fails.

5. Based on two different bus systems and 26 days of MTA data and over 14 days of SEPTA we recorded 97% facial tracking accuracy since the system knows when it can't track and stops measuring fatigue. We also reported 95% accuracy for distraction detection.

6. Our very encouraging results open the way for future further development of our system and the creation of a commercial product capable of detecting distraction, fatigue and driver identity, coupled with GPS and other vehicle data, as well as other vehicle cameras.

7. Use of our system has the potential to contribute to safer trucks, safer highways and improved transport of cargo.

Following the success of this initial pilot study, we will work with our program manager to communicate with truck manufacturing companies in order to potentially integrate our system into their driver cabin design. Our aim is to actively pursue several avenues to ensure the use and commercial future success of our system that will result in safer highways and reduction in accidents and loss of lives.

REFERENCES

[1] Milborrow S, Nicolls F. Locating facial features with an extended active shape model[C]//European conference on computer vision. Springer Berlin Heidelberg, 2008: 504-513.

[2] J. Huang, X. Huang, D. Metaxas, “Learning with dynamic group sparsity”, 2009.

[3] X. Yu, Z. Lin, J. Brandt and D.N. Metaxas, "Consensus of Regression for Occlusion-Robust Facial Feature Localization", 13th European Conference on Computer Vision, Sep., Zurich, 2014

[4] X. Yu, J. Huang, S. Zhang, W. Yan and D.N. Metaxas, "Pose-free Facial Landmark Fitting via Optimized Part Mixtures and Cascaded Deformable Shape Model", *14th IEEE International Conference on Computer Vision*, Dec., Sydney, 2013.

[5] P. Felzenszwalb, D. Huttenlocher, “Pictorial Structures for Object Recognition”, International Journal of Computer Vision, Vol. 61, No. 1, January 2005

[6] X. Yu, Z. Lin, S. Zhang and D.N. Metaxas, "Nonlinear Hierarchical Part-based Regression for Unconstrained Face Alignment", 25th International Joint Conference on Artificial Intelligence, 2016