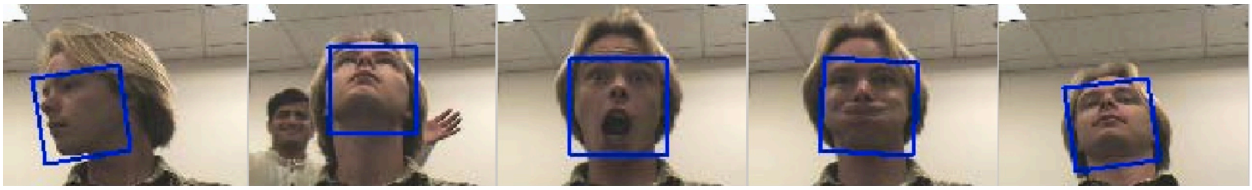
Design Review Report #1

VE/VM 450 Team 22

Face Tracking



Sponsor UAES

Instructor Chengbin Ma

Team Member Huan Zheng
Xuefeng Hu
Siying Li
Shengjie Pan
Zhenren Lu

Emails huanzheng1119@outlook.com,
leonhuxf@umich.edu,
{alicelsy, jessepsj, lzr950113}
@sjtu.edu.cn

The image is cited from Jiang, Yanxia, Hongren Zhou, and Zhongliang Jing. "Visual tracking and recognition based on robust locality preserving projection." Optical Engineering 46.4 (2007): 046401-046401.

1. Abstract

Our team is working with UAES to build an in-car object tracking system. The purpose of this system is to let the drivers interact with automotive electronics using simple gestures captured by a camera. However, due to the lack of data of hand gestures, we redirect our project slightly to track human face in the video. We will make our project highly portable so the project can be utilized to recognize gestures given enough data.

In this report, we will start with a detailed introduction to the problem in Section 2 and illustrate why we are building this system. In Section 3, we will summarize the materials we collected so far and introduce the baseline model as well as the benchmark for our project. The requirements and specifications are listed in Section 4, and we also show the QFD and explain how we reach it. Finally, we will provide the whole project plan in Section 5 and a Gantt chart to show the flow of our project. We will end this report with a short biography of each member in our team.

We also take into consideration the comments received upon our presentation in this report. For example, Prof Chengbin Ma asked about why we want to develop this project while there are many existing applications. We will integrate the latest theories and algorithms on Deep Learning and achieve a high accuracy compared to existing methods. And Prof Roberto Dugnani mentioned that the blink tracking system can be used as a benchmark against our project. We will collect relevant information and make the comparison in the next steps.

2. Problem Description & Introduction

Our sponsor UAES (United Automotive Electronic System) is looking for an innovative in-car system that helps the driver easily control automotive electronics without pressing any button and, thus, improve the driving experience. The most important feature of the system is to recognize different gestures using a camera in the vehicle. However, this is a completely new project without existing framework or testing data. As our mentor suggested, we start to build a highly portable project and make necessary adjustments in the future. So we decide to use deep learning as our method.

The sponsor initially requested a gesture tracking system, which requires a large quantity of labeled data about hand gesture to train and test our project. It will take us more than one month to collect and label all the data, and we will have limited time to improve our system. After consulting the sponsor, our team decided to redefine the problem as an object tracking system, and, specifically, face tracking. Compared with gesture tracking, face tracking has more mature open-source database and API. We will build a face tracking system with well-defined data as well as the existing framework. To meet the sponsor's needs, we will build a highly portable system, which can be easily adapted and transformed to gesture tracking given enough data.

3. Information Sources

3.1 Literature study summary

We used Google scholar and made a throughout research on the face tracking and object tracking topic. We found that there are basically two directions to build a tracker:

- Target representation and localization: A bottom up method, which aims on using the detector to detect the object in each frame and make them together to be the track series. Representative algorithms: mean-shift tracking [4], contour tracking [5]
- Filtering and data association: A top-down method, which aims on using the information of the previous frames to predict the location of the object in the next frame. This method often have higher accuracy and robustness, but also often acquire higher computational complexity. Representative algorithms: Kalman filter [6], Particle filter [7].

3.2 Baseline Model

The baseline model we are going to use is called the Kernelized Correlation Filter [1]. The work is published in 2015 and has 583 citations until now. It the most popular tracker people uses after 2015. The KCF is basically an online learning model using kernelized regression. However, it takes advantages of the structure of the video frames and image patches, and translates the solution of the regression model into Fourier space to avoid the calculation of the matrix inversion, which cost the most time in classic algorithms. The KCF outperform all the other trackers at the time under the 50 videos benchmark [2], and we are going to build our model based on it.

3.3 Benchmark

The benchmark we are going to use is a video database published in 2013. The database contains 50 videos, including several groups of videos with specialized constrained situations. This benchmark is very suitable for video algorithm testing, and it is the same benchmark that the KCF paper used.

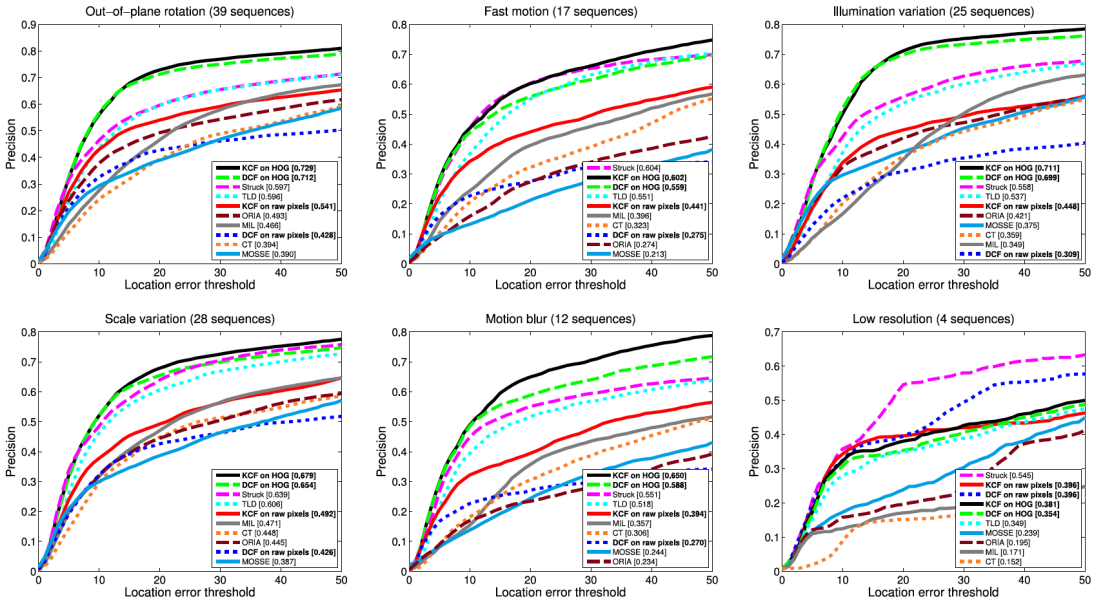


Fig 1. Baselines: the performance of KCF under 6 special situations [1]

4. Customer Requirements & Engineering Specifications

4.1 Customer Requirement

Our customer, UAES, is looking for a fast accurate reliable tracker that can be installed in a car. However, the proposal from them is too general that they not even define the object of the tracker. After an in-depth discussion with the mentor, we summarize requirements from customers as the following:

- Object: gestures/face/general objects in a car
- Real-time
- Accurate
- Reliable under different cases
- Portable/being able to be installed in a car

4.2 Engineering Specifications

To translate the customer requirement into engineering specification, we first talked to the professor who teaches Computer Vision, he shared his experience with us that data collection is the most time-consuming part. He also mentioned that there is a huge amount of datasets on the face with great diversity. Considering the limited time, we choose human face as our tracking object. This decision will ease our burden on collecting and labeling data.

Secondly, we did the literature research in order to find the benchmark and leading algorithms in the tracking area. Through reading the papers, we also learn how the performance of an algorithm is evaluated from academic aspects. We find that

- Real-time = Fast speed/low time delay: Frame Rate (>24 frames per second (fps))
- Accurate = high precision: Average Precision (AP) (%)

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

- Reliable under different cases = Robustness: Precision Variance/Recall (%)

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

- Portable/being able to be installed in a car = size of the model/parameters (M)

4.3 QFD

Quality Function Development (QFD)

Correlation between eng. Specs

++ Strong Positive
+ Medium Positive
- Medium Negative
- - Strong Negative

	Weight (1-10)	Network's layer amount	Parameters size	Train time	Computing Cost	Frame Rate	Resolution of input	Recall	Batch size	Precision variance under 6 cases	Mean Average Precision	Benchmarks: Kernelized Correlation Filter tracker		
Speed	10	3	3		1	9	9					3		
Precision	8		3	3			3	3		1	9	1		
Reliable/Recall	6							9		3	3	5		
Portability	5	9	9		9							1		
High robustness	8			1			1		3	9		4		
Stability of Training	5	1		9	3				9	3		2		
Measurement Unit		num	M	epoch	USD	FPS	P	%	num		%			
Target Value		19	100	10	300	24	480	80	128	10	80			
Total		80	99	77	70	90	122	78	69	113	90			
Normalized		0.09	0.11	0.08	0.08	0.10	0.13	0.09	0.08	0.12	0.10			

Competitive benchmarks
5 = satisfies perfectly
4 = satisfies "mostly"
3 = satisfies "somewhat"
2 = satisfies "slightly"
1 = doesn't satisfy at all

Correlation between customer reqs and eng. Specs
9 = Strong Relationship
3 = Medium Relationship
1 = Small Relationship
(blank) = Not Related

Fig 2. QFD

As the requirement are listed above, we find the relative factors may affect the performance including:

- Networks' layer amount: The more layers, the stronger learning ability it has. However, the speed will decrease with increasing amount.
- Parameters' size: The larger it is, the slower the model performs on PC, and while the stronger learning ability it has.
- Training time/computing cost/batch size/resolution of input: it will affect the portability also precision.
- Frame Rate/Frames per second: This should be in the tradeoff with precision.
- MAP/Precision Variances/Missing Frames: it directly measures the performance of the tracking system.

The following table contains all the engineering requirements of our project.

Specifications	**Target Value**
The amount of networks' layer	19
The size of parameters	100 (MB)
Training time	10 (Epoch)
Computing cost	300 (USD)
Frames per second	24 (FPS)
Resolution of input	480 (P)
Recall	80%
Batch size	128
Precision variance	10
MAP	80%

Table 1. Engineering Requirements

5. Project Plan

5.1 Overview

Our project plan is closely following the schedule of the course, which includes three design reviews and a final expo. Besides the course requirements, we also consider the most time-consuming part of our project, the training of our models as our milestones. We schedule three training in total, each training period includes training, result checking, and improving the results. After the first training, we will determine our baseline. After the second training, we will fix the model and adjust the parameters. After the third training, we will finish the model and create a demo version. We are currently in the preparation stage, during which we should study the relevant literatures, prepare the training and testing environment, and collect plenty data for our project. Besides the coding part, we also take into consideration the preparation of design review and reports.

5.2 Team Assignment

Huan Zhen is the team leader. She will be in charge of data collection and monitor the entire project. Xuefeng Hu is the researcher. He will study relevant papers and algorithms. Shengjie Pan and Zhenren Lu are the engineers in the team. They will work on the coding part as well as debugging. Siying Li is the writer. She will mainly prepare the reports and presentations. The assignment is not strict. All teammate will collaborate with each other to finish the tasks.

5.3 Budget

The budget of the project is mainly composed of the rent cost of AWS services. We need approximately two months to finish the project. The total budge is ¥3,200, with ¥3,000 server cost and ¥200 other necessary costs, e.g. transportation, printing.

⊕	Amazon EC2 Service (US-East)	$	238.46
⊕	Amazon S3 Service (US-East)	$	0.91
⊕	AWS Data Transfer In	$	0.00
⊕	AWS Data Transfer Out	$	0.81
⊕	Data Transfer Out to CloudFront	$	0.00
⊕	AWS Support (Basic)	$	0.00
Free Tier Discount:		$	-2.98
Total Monthly Payment:		$	237.20

Fig 3. AWS Service Cost

5.4 Gantt Chart

	Name	Duration	Start	Finish
1	Server Selection	3d	06/07/2017	06/09/2017
2	Preparation for DR1 presentation	4d	06/07/2017	06/12/2017
3	Framework Research	7d	06/07/2017	06/15/2017
4	Open Source Code Selection	7d	06/07/2017	06/15/2017
5	Data Preparation	10d	06/07/2017	06/20/2017
6	DR1 Report	6d	06/13/2017	06/20/2017
7	Image processing	4d	06/21/2017	06/26/2017
8	Data Interface Modification	4d	06/16/2017	06/21/2017
9	Pre-trained Model Selection	3d	06/21/2017	06/23/2017
10	Preparation for DR2 Presentation	4d	06/23/2017	06/28/2017
11	First Training (Baseline)	5d	06/27/2017	07/03/2017
12	Plot the Result of the Baseline and Check	2d	07/04/2017	07/05/2017
13	Read Papers and Seek Ideas to Improve	5d	06/27/2017	07/03/2017
14	Coding Based on the Idea	7d	06/30/2017	07/10/2017
15	Second Training	5d	07/11/2017	07/17/2017
16	Plot the Result and Check	1d	07/18/2017	07/18/2017
17	Read Papers and Seek Ideas to Improve for Final Cases	4d	07/13/2017	07/18/2017
18	Adjust Parameters for the Final Cases	4d	07/14/2017	07/19/2017
19	Preparation for DR3 Presentation	4d	07/14/2017	07/19/2017
20	DR3 Report	5d	07/20/2017	07/26/2017
21	Third Training	5d	07/20/2017	07/26/2017
22	Plot the Result and Adjust for the demo	4d	07/27/2017	08/01/2017
23	Coding for Demo	15d	06/15/2017	07/05/2017
24	Final Report	5d	08/01/2017	08/07/2017
25	Design Expo Preparation	3d	08/07/2017	08/09/2017

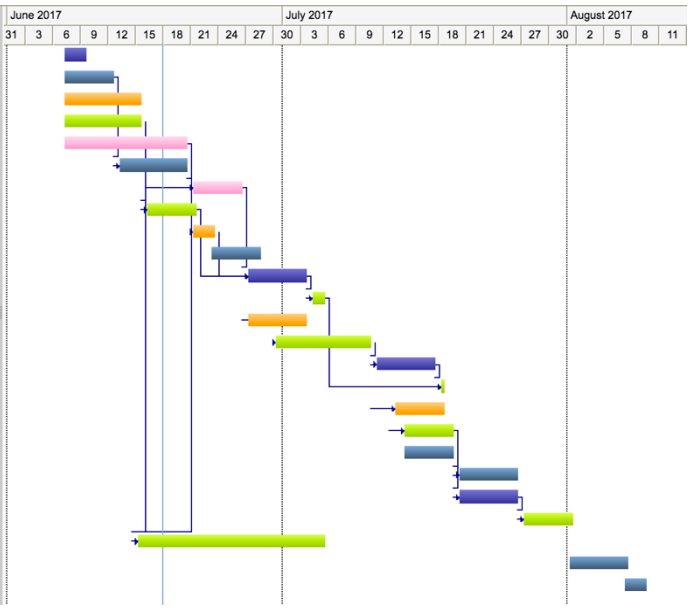


Fig 4. Gantt Chart

In the Gantt chart, the green bar represents the coding part, the orange bar represents the literature study, the pink bar represents data collection, the purple bar represents three trainings, and the blue bar represents the preparation for design reviews and final expo. The three milestones are 6/27, 7/11, and 7/20.

6. Conclusion

In this report, we summarize the progress so far and give an overview of our project. We define the problem, provide a summary of the literature study, list specifications and requirements, and show the detailed plan of our project. We are currently up to the schedule and anticipate to accomplish the rest of the project.

7. References

[1] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596.

[2] Y. Wu, J. Lim, and M. H. Yang, “Online object tracking: A benchmark,” in Proc. Comput. Vis. Pattern Recognit., 2013, pp. 2411–2418.

[3] Jiang, Yanxia, Hongren Zhou, and Zhongliang Jing. "Visual tracking and recognition based on robust locality preserving projection." Optical Engineering 46.4 (2007): 046401-046401.

[4] Comaniciu, Dorin, Visvanathan Ramesh, and Peter Meer. "Kernel-based object tracking." IEEE Transactions on pattern analysis and machine intelligence 25.5 (2003): 564-577.

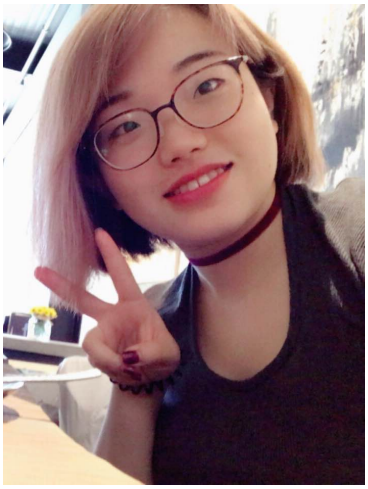
[5] Baumberg, A. M., and D. C. Hogg. "An efficient method for contour tracking using active shape models." Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on. IEEE, 1994.

[6] Qian, Richard J., M. Ibrahim Sezan, and Kristine E. Matthews. "A robust real-time face tracking algorithm." *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. Vol. 1. IEEE, 1998.

[7] Nummiaro, Katja, Esther Koller-Meier, and Luc Van Gool. "An adaptive color-based particle filter." Image and vision computing 21.1 (2003): 99-110.

8. Bios

Huang Zhen



I am a senior student in ECE major named Huan. I have been engaged in Computer Vision specially Object Detection for over one year. I am currently on a research internship for Computer Vision on Face Detection which is relevant with this topic offered by Intel, Inc. I am working as researcher in the Algorithms Platform (算法平台) at Horizon Robotics in Beijing.

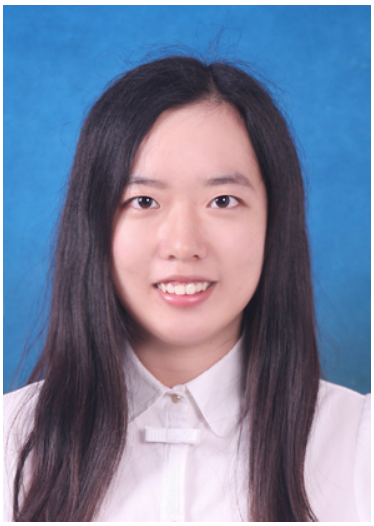
Xuefeng Hu

Academic Background: A Electronic and Computer Engineering student at Shanghai Jiao Tong University, UM-SJTU Joint Institute and a Computer Science student from University of Michigan.

Researched Interests: I have research experiences in computer vision with Professor Jia Deng and quantum computing with Professor Yaoyun Shi. Now I am master student in Computer Science at the Columbia University, and I may join Professor Shih-Fu Chang's lab this Fall. My major research interest is to build an intelligent system that understands video events (or even movies). I am going to continue pursuing the Ph.D. degree after my master at Columbia.

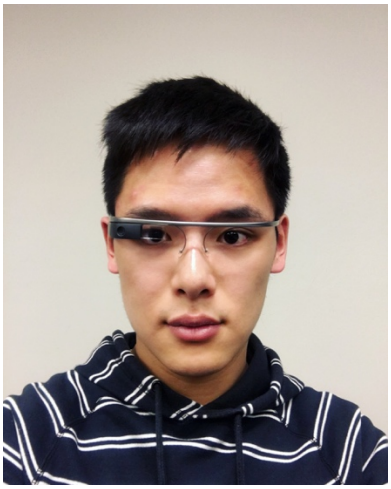
Other interests: A huge fan of Leonard Cohen, William Somerset Maugham, Raymond Carver, Kevin Spacey and cats. Love playing guitar, writing and sketching. Enjoy live-house rock music show a lot. A weird youth who still read poems.

Siying Li



I'm a graduating senior ECE student at Shanghai Jiao Tong University and has graduated from University of Michigan majored in Computer Engineering. I mainly focused on Embedded system and Robotics for undergraduate's studies and have some experiences on Computer Vision. I had a project for matching human faces and animal faces in the Computer Vision course in UM with my teammates. I will keep studying in Computer Vision and Machine Learning area for Master degree in University of Michigan. In my previous internship, I worked for Hinacom in Beijing researching on medical image detection in summer of 2016.

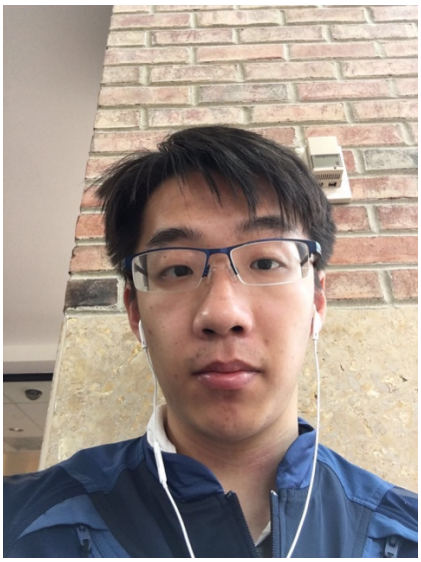
Shengjie Pan



Hi, I'm Shengjie Pan. I am a DD student studying CS at UM and ECE at JI. I find myself 100% CS guy, because I'd rather spend all day debugging and coding than working in the lab. I have experience in machine learning and data mining. The advanced courses I took at UM, such as database system, machine learning, and information retrieval, largely expanded my horizon in CS. I have also researched on graph mining and HCI, and I'd like to try more topics. That's also

part of the reason why I choose this project as capstone. For the future, I think I will end up in the industry as a software developer or “Ma Nong”. But before that, I will take a master program to acquire state-of-art knowledge.

Zhenren Lu



Zhenren Lu is a CS majored Bachelor graduated from UM and undergraduate student of SJTU majored in ECE. During his undergraduate career, his major interests is Machine Learning. He has taken the course EECS445(Intro to Machine Learning), EECS492(Intro to Artificial Intelligence) and the Stanford online CV course CS231(Convolutional Neural Network for Visual Recognition). He also has decided to pursue a master degree in the field of ML and AI after his graduation from SJTU.