DESIGN AND IMPLEMENTATION OF INTEL-SPONSORED REAL-TIME MULTI-VIEW FACE DETECTION SYSTEM

Bohua Gan[1], Vincent Chang[1], Guanying Wang[1], Xiuli Pan[1], Guan Wang[1], Naihai Zou[2], and Felming Feng[2]

[1]University of Michigan—Shanghai Jiao Tong University Joint Institute, Shanghai, China
ganbohua@gmail.com
[2]Intel Asia-Pacific Research and Development Ltd, Shanghai, China
nanhai.zou@intel.com

ABSTRACT

The paper introduces a case study of design and implementation of Intel-sponsored real-time face detection system conducted in University of Michigan—Shanghai Jiao Tong University Joint Institute (JI). This work is teamed up totally by 15 JI students and developed in three phases during 2013 and 2014. The system design of face detection is based on Intel High Definition (HD) 4000 graphics and OpenCL. With numerous techniques including the accelerated pipeline over CPU and GPU, image decomposition, two-dimensional (2D) task allocation, and the combination of Viola-Jones algorithm and continuously adaptive mean-shift (Camshift) algorithm, the speed reaches 32 fps for real-time multi-view face detection. Plus, the frontal view detection accuracy obtains 81% in Phase I and reaches 95% for multi-view detection, in Phase III. Furthermore, an innovative application called face-detection game controller (FDGC) is developed. At the time of this writing, the technology has been implemented in wearable devices and mobile with Intel cores.

KEYWORDS

Multi-view face detection, real-time, graphic processing unit (GPU), OpenCL, face-detection game controller (FDGC).

1. INTRODUCTION

Mobile commerce has reached unprecedented height: global smart phone shipments have over 1.16 billion in 2014. Meanwhile the competition of mobile processor chip market is fierce between Qualcomm, MediaTek (MTK), Samsung, and Intel, etc. Qualcomm and MTK achieve above 30 percent of market share each, following by Samsung's 20 percent of the share. Unlike its predominant advantage in the PC fields, Intel gains only 2.81 percent share in 2014 which nearly missed the mobile revolution [1].

Intel's core competitiveness hinges up advanced semiconductor fabrication process and technology. In addition, effective and powerful software solutions make Intel the leading position in the semiconductor industry. Furthermore, Intel has various System on Chip (SoC) for different platforms and billions expenditures of Research & Development every year. Advanced technologies such as SoC fabrication, Bluetooth, and wireless charging etc. are widely used in Intel-inside wearable and mobile devices.

RealSense technology which incorporates face recognition and three-dimensional (3D) cameras is used on the robots, unmanned aerial vehicles, as well as a new interactive mode of 3D immersive experience. This innovative mode could eventually make user free from the constraint of keyboards and passwords.

As the most important part of RealSense technology, face detection has been in-depth studied by Intel. This corporation has sponsored   three-phase real-time face detection capstone design project (CDP) on University of Michigan－Shanghai Jiao-Tong University (Joint Institute) including 15 students during 2013 and 2014 [2]. It aims at innovating in the face detection technology and improving the performance of detecting speed, and accuracy etc.

The speed of original version which at start of Phase I is 3 fps for 480p videos frontal view face detection and it reaches 9 fps at the end of Phase I. With the combination of image decomposition and two-dimensional (2D) task allocation in Phase II, the speed is increased by 3 times and reaching 30 fps. Combining the Camshift tracking in Phase III, multi-view face detection speed is 32 fps finally. Plus, the detection accuracy obtains 81% in Phase I and reaches 95%, in Phase III. In addition, an interesting interactive application called Face-Detection Game Controller (FDGC) is developed. The program's conception, design details and testing results are presented in Section 2. Conclusions are given in Section 3.

2. DESIGN AND IMPLEMENTATION OF THE REAL-TIME MULTI-VIEW FACE DETECTION SYSTEM

As Table 1 shows, there are three phases in the face detection CDP, including two phases in 2013 and one phase in 2014. Every phase lasts about 12 weeks and involves 5 students.

2.1. OpenCL Accelerated Face Detection (Phase I, summer, 2013)

The first step is a literature survey and an objective analysis. According to the capability of Intel's HD 4000 graphics and OpenCL, the customer's requirements are definite; they are performance improvement, working prototype, and deliverable product separately. To meet the customer requirements, the detailed engineering specifications are generated in Week 1. There are mainly three aspects which are accuracy, efficiency, and innovative application [3]. As Table 2 shows, the objective of efficiency is reaching over 10fps for 480p videos, detecting 85% frontal view faces correctly, and implementing an innovative application with bulged faces detection.

In Week 2, the entire video processing flows are definite as shown in Figure 1.The entire processing function is decomposed into three workflows: video pre-processing, face detection and accelerated process &innovative apps.

- Video Processing comprises of a series of preprocessing on the input video. It changes the video stream into a different type of data which will make subsequent computation easier.
- Face detection deals with the computation of images to locate human faces and find the accurate sizes of them.
- Accelerated process will improve the speed exponentially and innovative applications are concerned with how to implement application programs after face detection is done.

Through brainstorming, in which every member expressed their ideas of possible solution concept related to the projects and Quality Function Deployment according to the discussion, the combination of using Segment into Shots, Viola-Jones Algorithm, and GPU acceleration was selected as the final design.

Table 1. Overview of all the tasks involved in developing this 3-phase work during 2013 and 2014. The task marked with * represents the primary contribution for acceleration.

Week(nth)	Phase I (Summer, 2013)	Phase II (Fall, 2013)	Phase III (Summer, 2014)
1	Literature Survey, Objective Analysis		
2	Function Decomposition &Algorithm Choosing	Previous Work Rebuilding	
3	Image Face Detection		
4	Video Stream Face Detection	Analysis & Design	Skin-Color-Based Image Segmentation Implementation*
5			
6	Performance Improvement	Pipeline of Accelerated Concept over CPU and GPU & Image Decomposition and Task Allocation*	
7	Innovative Applications Design		Viola-Jones detection + Camshift tracking Pipeline Building*
8			
9	Debug & Test		Tuning Face Recognition using Principal Components Analysis*
10			
11	Final Delivery	UI Design, User Feedback and Update	Detection Box Merge Optimization & Image Preprocessing Pipeline Tuning*
12		Final Delivery	

Table 2. Engineering specifications of Phase I project, OpenCL acceleration face detection

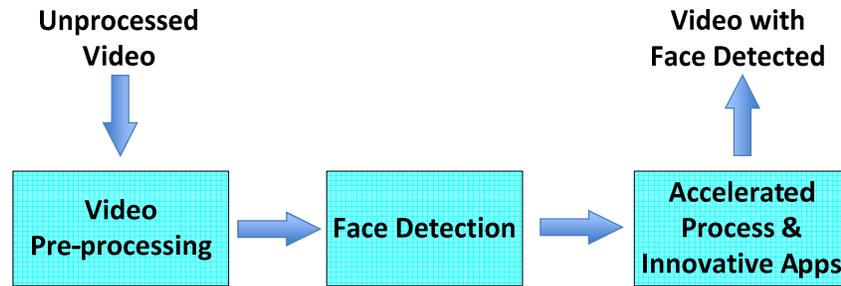| Engineering Specifications	Description
Accuracy	Detect 85% faces correctly
Efficiency	10-15 fps for 480p videos with frontal view face detection
Innovative Application	Adding bulged effect on the detected faces

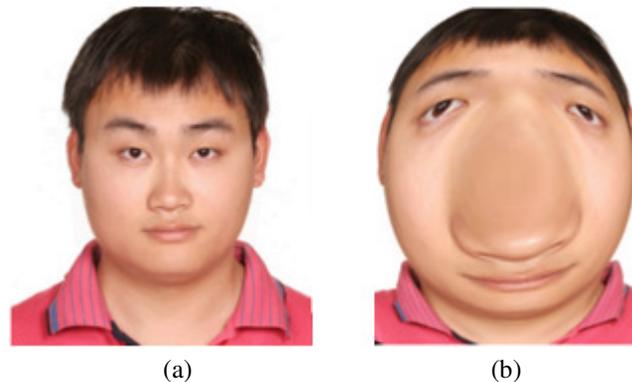Fig. 1. Video processing flow in Phase I

Fig. 2. A special effect on the detected face in Phase I: (a) Original image (b) The image added with the bulged effect on the detected face (Courtesy of Ref. [3]).

- Segment into shots, images are extracted on a certain rate and the number of extracted images is reduced.
- Viola-Jones algorithm, it provides fast and efficient face detection framework [4].
- GPU acceleration, images separated from video streams can be dealt in parallel with GPU. Therefore the speed is increased

From Week 3 to Week 5, the image and video stream face detection are completed. Due to slow data exchange between CPU and GPU and the many sub-images that needed to be processed, the performance of the initial version running in Intel HD 4000 graphics did not meet the expectations. In order to improve the speed and accuracy, processing all sub-images in parallel, exchanging data only when necessary, and buffering small exchanges are applied. Therefore, the frontal view face detection speed has changed from 3 fps to 9 fps for 480p videos which is little less than expectation. The accuracy is 81% in average.

Innovative application design is an important part of this project. And according to the mentor from Intel, students are encouraged to think outside of the rules. The team implements bulge effects on the detected faces as shown in Figure 2, and faces in the video are all bulged smoothly. In Week 9 and 10, the team focused on testing and debugging. They found several bugs when running with Intel hardware. This helps Intel improve the performance of the hardware, and it is also beneficial to the following capstone design teams.

2.2. OpenCL Accelerated Face Detection (Phase II, fall, 2013)

Intel requires the team of Phase II to bring the Phase I face detection program to a new level with the help of Intel Beignet OpenCL implementation. Specifically, two specific requirements were mentioned by Intel:

- Optimize the existing parallel computing algorithm.
- Come up with innovative ideas on applications and realize some of them [5].

Based on first three weeks' literature survey, objective analysis, and previous work rebuilding, the engineering specifications are definite as shown in Table III. The goal of this phase is faster, more accurate and innovative.

The pipeline of accelerated concept over CPU and GPU has displayed in Figure 3 and was decided particular by the team in Weeks 4 and 5. The general process is to first turn video streams into still frame which will divert into gray images in CPU. Classifier will get the image faces data from gray images, which include information of the size and location, and then add special effects on the faces in GPU. Thus making GPU and CPU work simultaneously, the speed increases obviously and reaches 20 fps for 480p videos.

In order to further accelerate the process, some ideas were considered and two of them were used. They proved to be available in Phase I. In this phase, the extensive improved solutions, including image decomposition and task allocation, are designed in detail. Image decomposition is to make pre-processing and classifiers more efficient, and task allocation is to make the acceleration over CPU and GPU more efficient [6]. Due to the effort in these two factors, it reaches 30fps for 480p videos which is higher than Phase I and the accuracy achieves 94% in average.

(1) Image decomposition. Since a GPU has the property of parallel calculation, the images were separated into small parts and then transformed together into the GPU. With the little parts calculated at the same time, the calculated speed is raised remarkably.

(2) Task allocation. The tasks are sorted in a chain before, but the cores of GPUs are distributed like a 2D matrix. It takes quite a lot of time to match up each task with each core of a GPU. In this phase, tasks are put into a 2D task matrix where each task is matched up with the cores in a GPU automatically, and the matching time is greatly reduced.

Innovative application is the vital engineering specification in Phase II, so the team pays much more attention on user interface and creative application design. In the final version of the program, it supports two types of inputs, which are still images and video streams. As shown in Figure 4, users can click the "Video" button or "Image" button to choose different input forms. For choosing video as input, users need to make sure that the computer has been connected to a web camera firstly. The program will do real-time face detection on the videos

captured by the camera. For choosing image as input, users need to click the "Browse" button as the input document.

Table 3. Engineering Specifications in Phase II.

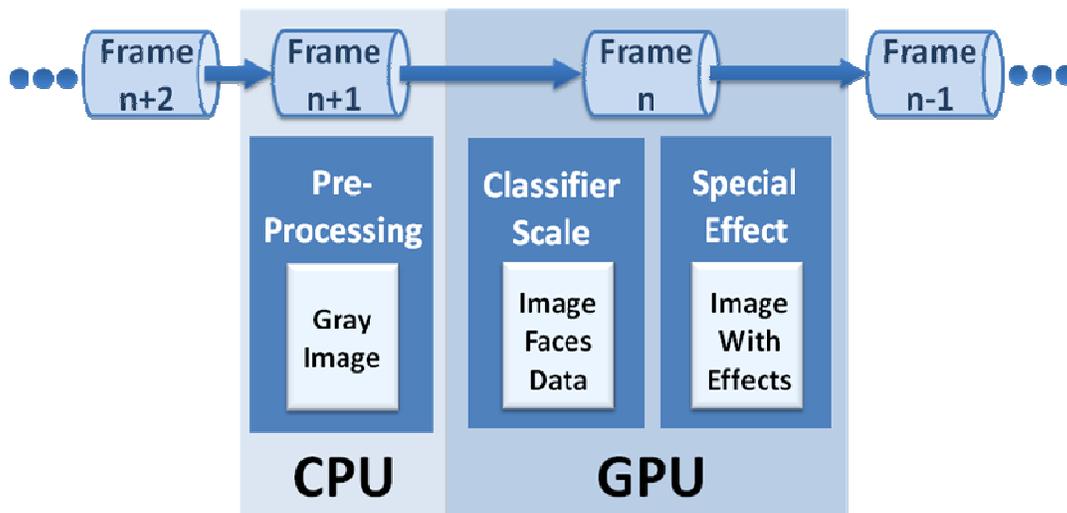| Engineering Specifications	Description
Accuracy	Detect 90%-95% faces correctly
Efficiency	20-30 fps for 480p videos with frontal view faces detection
Innovative Application	Adding different effects on the detected faces

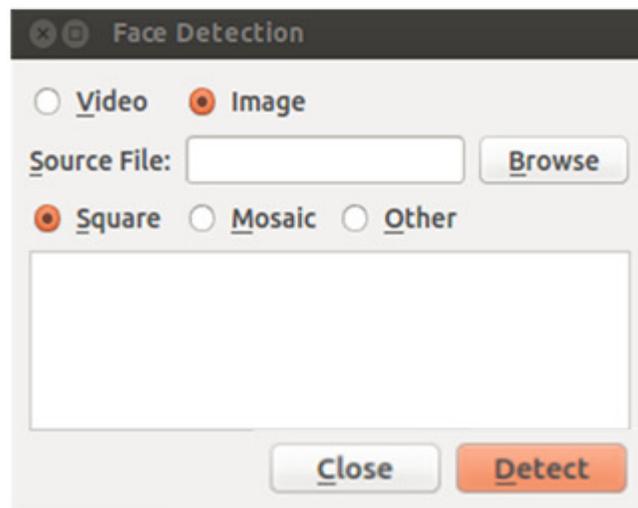Fig. 3. Design of the pipeline for face detection acceleration over CPU and GPU in Phase II.

Fig. 4. User Interface of the accelerated face detection system in Phase II (Courtesy of Ref. [6]).

(a)   (b)   (c)

Fig. 5. Specialeffects on the detected faces in Phase II: (a) Rectangle (b) Mosaic (c) Comic relief (Courtesy of Ref. [6]).

In addition, users can click three options for face displaying as shown in Figure 5. They are the rectangle, mosaic and comic relief effects.

2.3 OpenCL based Face Detection and Recognition Algorithms Tuning (Phase III, summer, 2014)

Different from previous phases, Phase III is asked to realize the tilted face detection, and detecting multi-view faces makes the product satisfactory with the commercial requirements as shown in Table 4 [7].

Three image pre-processing steps which have shown in Figure 6 are used to remove background. Detection area is significantly reduced, and the speed is boosted with these techniques.

- Skin Segmentation is used to identify skin region. Firstly, RGB image is converted to YCrCb image which ignores illumination, then skin model is applied to obtain segmented binary image. Meanwhile, some topological methods, like dilation, erosion and Gaussian smooth, are used to denoise and smooth the binary image [8].
- Bounding Boxes Finding is used to find the located detection area. Label the connected white components in the binary image at first and find contours of each component. Then rectangle windows are drawn to bounding each contour.
- Merge the overlapped boxes into larger joint boxes. The merge operation can reduce duplication caused by the overlap which will reduce the execution time.

Table 4. Engineering Specifications in Phase III.

Engineering Specifications	Description
Accuracy	Detect 90%-95% faces correctly
Efficiency	Over 30 fps for 480p videos with multi-views detection
Functionality	Tilted face detection
Innovative Application	Face-Detection Game Controller

From Week 7 to Week 10, the team focused on increasing the accuracy, solving the problem of detecting tilted faces and video face detection which based on the image face detection. Camshift (Continuously Adaptive Mean-shift) [9] can detect tilted face in a large scale of angle and the

inner core of it involves more float and matrix calculation than Viola-Jones. The solution which builds a pipeline that leverages the consistency of video with Viola-Jones and Camshift algorithm. The new pipeline for the video multi-view face detection shows in Fig. 7 and there are three steps as follows:

- Using Viola-Jones Algorithm to detect faces every 10 frame.
- Using Camshift Algorithm to track those faces for 9 continuous frames.
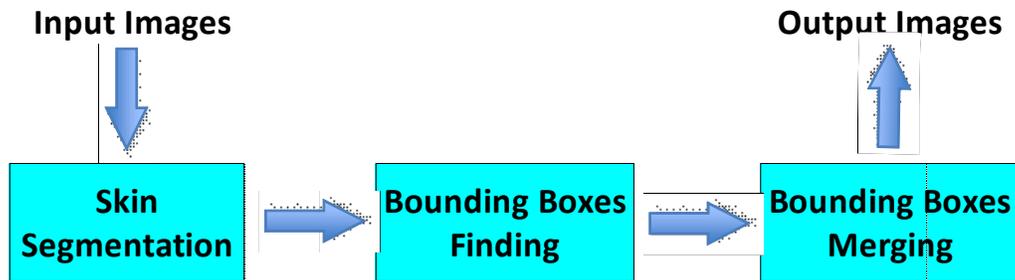- Iterate previous two steps.

Fig. 6. Flow chart for image pre-processing in Phase III (Courtesy of Ref. [7]).
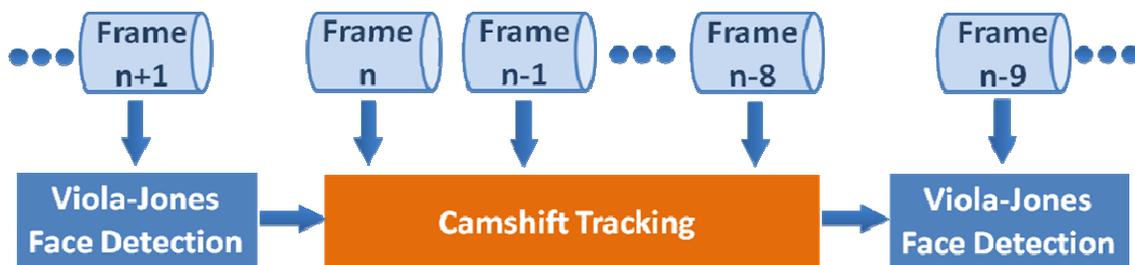

Fig 7. New pipeline combining Viola-Jones Face Detection with Camshift Tracking (Courtesy of Ref. [7]).

Fig. 8. An innovative application in Phase III: A face-detection game controller (FDGC)
(Courtesy of Ref. [7]).

According this new pipeline, the speed breaks through to 32 fps for multi-views face detection and the accuracy is over 95%. In order to show the improvement of Phase III and get real value with new pipeline technic, the team applies it in Face-Detection Game Controller, in which everyone can throw the keyboard, mouse away and play the video game by the face as shown in Figure 8. Four more steps are needed for this extended pipeline:

- Detect & Track Faces, using pipeline to detect and track faces.
- Get the Angle of Faces, using geometrical principle to calculate the angle.
- Socket Communication between Windows & Unix, using socket communication to control signals (i.e. Left/Right).
- Simulate Keyboard Signal, using windows inner function to simulate hardware signal and operate the game.

3. CONCLUSIONS

Intel sponsored 3-phase real-time face detection capstone design projects executed at JI during 2013—2014 in order to strengthen its RealSence technology. The ultimate goal of this work is to catch up the market demands for mobiles and wearables as well as compete with predominant industry giants. Prior to Phase I, the speed of the original version is 3 fps for 480p videos frontal view face detection.

The platform development and system design of this work are based on Intel's HD 4000 graphics and OpenCL. It reaches 9 fps and 81% accuracy after the platform setup and system validation in Phase I. With the techniques of image decomposition and 2D task allocation, the speed and accuracy increase to 30 fps and 94%, respectively, for frontal view detection in Phase II. Combining Viola-Jones algorithm with the Camshift tracking method, the speed reaches 32 fps and the accuracy achieves 95% for real-time multi-view face detection in Phase III. Furthermore, an interesting interactive application called face-detection game controller (FDGC) is developed. The technology has been implemented in wearable devices and mobile with Intel cores at the time of this writing.

REFERENCES

[1] R. James(2014), Brainstorm Tech conference in Aspen, Colorado. Available: http://fortune.com.
[2] V. Chang (2014), "MNC-Sponsored Multidisciplinary Industrial-Strength Capstone Design Projects in China," International Journal for Innovation Education and Research, Vol.2-10, pp 54-65.
[3] G. Wang, H. Chen, G. Li, Z. Fu, and X. Guo (2013), "OpenCL Accelerated Face Detection (Phase I)," undergraduate thesis, University of Michigan—Shanghai JiaoTong University Joint Institute, summer.

[4]   P. Viola, and Michael J. Jones (2004), "Robust Real-Time Face Detection," International Journal of Computer Vision, Vol. 57, pp 137-154.

[5]   X. Pan, S. Cao, Z. Chen, J. Zhu, and X. Liu (2013), "OpenCL Accelerated Face Detection (Phase II)," undergraduate thesis, University of Michigan—Shanghai Jiao Tong University Joint Institute, fall.

[6]   C. Fernandez, J. R. Saeta, X. Martorell, and J. Hernando (2011), "Real-time GPU-based Face Detection in HD Video Sequences," IEEE International Conference on Computer Vision Workshops, November.

[7]   G. Wang, L. Su, B. Zhang, S. Zhu, and B. Hong (2014), "OpenCL Based Face Detection and Recognition Algorithm Tuning (Phase III)," undergraduate thesis, University of Michigan—Shanghai Jiao Tong University Joint Institute, summer.

[8]   H. Han, S. Shan, X. Chena, and W. Gao (2013), "A comparative study on illumination preprocessing in face recognition," Pattern Recognition, Vol. 46, pp 1691–1699.

[9]   C. Du, H. Zhu, L. Luo, J. Liu, and X. Huang (2013), "Face detection in video based on AdaBoost algorithm and skin model," The Journal of China Universities of Posts and Telecommunications, Volume 20, Supplement 1, pp 6–9, 24.

AUTHORS

Bohua Gan is a postgraduate student of Shanghai Jiao Tong University majors in Electrical and Communication Engineering. He received B. E in Communication Engineering from Dalian Maritime University in 2013 and LL. B in Law from Dalian Maritime University in 2012. He is interested in all kinds of techniques, such as open source GPGPU, antenna, and telemedicine etc.

He is also a research assistant of Prof. Vincent Chang since September, 2014. He made effort in Intel-sponsored Capstone Design Projects focusing on OpenCL related tasks on Intel GPU. He wishes to make our lives better through the technology.

Vincent Chang, Ph.D. & MBA is Teaching Professor & Faculty Director of Corporate Relations at University of Michigan—Shanghai Jiao Tong University Joint Institute (JI). He is also the founder of Knowledge Master, Inc. (KMI) in the US and the cofounder of Semiconductor Leadership programs at UC Berkeley Extension.

Dr. Chang has been working in academia and industrial positions for more than 20 years in total with experiences in China, US and Taiwan. Dr. Chang pioneered Corporate Partners Programs at JI, covering entrepreneurial innovation and industry research. He strategically initiated and developed 96 MNC-sponsored Capstone Design Projects focusing on healthcare, energy, electronics, and mobile internet, creating an innovative platform collaborative with Covidien, GE, HP, Intel, Philips, and Siemens.

From 2004 to 2010, Dr. Chang was the President and CEO of KMI, a California-based e-learning company specializing in semiconductor IC design and innovation leadership. The company was recognized as an outstanding on-demand technology and content provider by UC Berkeley and awarded 9-year business partner contract from 2007 to 2016. Prior to moving to the US in 2003, Dr. Chang was Associate Professor of Electrical Engineering at Tamkang University in Taiwan. In 1995, he cofounded Calvin Engineering specializing in structural design along with Calvin Chang in Taiwan in 1995. His research has appeared in Electronics Letters, IEEE Journal of Lightwave Technology, IEEE Photonics Technology Letters, and International Journal for Innovation Education and Research. He also authored 14 textbooks in microelectronics published in Taiwan.

Guanying Wang is a software engineer at Google. She received B. E in Electrical and Computer Engineering from Shanghai Jiao Tong University, Shanghai, China and B.E in Computer Software Engineering from University of Michigan, Ann Arbor, in 2013. She received M. E in Information Technology from Carnegie Mellon University in 2014. She is interested in all kinds of novel technologies. She had participated in Mobile Advertisement System Design, Dynamic Keyboard IOS Application, Contextual Design, Inquires and User Interface Prototyping, Intel Innovation Project: OpenCL Accelerated Face Detection and Carnegie Financial Service Web App Development. And she is also a good manager of the project.

Xiuli Pan is a student of Shanghai Jiao Tong University majors in Electrical and Computer Science Engineering. He is interested in machines as well as programs for the logic inside them is attractive.
He is interested in all kinds of techniques such as network, microprocessor, automatic control and programming. With all of the skills I have learned I have made a PRP system based on microprocessor and Bluetooth to detect the barrier and guide the road on smart phones for blind people.
He always dream of making life better by making the techniques more easy to use and more ways to be used. And this idea makes me to join this project and want to make the face detection more easily to use and have a wild use in all areas.

Guan Wang was born in Beijing, China, in1992. He received B.S.E. degree in electrical and computer science engineering from Shanghai Jiao Tong University, Shanghai, China and B.S.E. degree in computer science engineering from University of Michigan, Ann Arbor, in 2014.
His research interests are are mainly in the deep learning, data mining, and information retrieval. His is doing research study with Prof. Honglak Lee. Try to build a system that detecting vehicle as well as road conditions by analyzing real-time video and pipelining multiple detection and tracking algorithms in machine learning and computer vision, such as Convolution Neural Network (CNN), and Discriminative Part Models (DPM). He is also a research assistant of Prof. Michael Cafarella. He made effort in Senbazuru, which is a prototype spreadsheet database management system (SSDBMS), and also helped developing a Feature Engineering Infrastructure that could dramatically ease the Explore-Extract-Evaluate interaction loop that characterizes many trained system projects. As for his future plan, he wishes to continue digging into deep learning and information retrieval area and leveraging those knowledge to make something meaningful.

Nanhai Zou is a senior staff engineer in Intel Asia-Pacific research and developing center. He works for Intel Open source technology center. He is leading a team to work on open source GPGPU and OpenCL related tasks on Intel GPU. Nanhai Zou graduated from Shanghai Jiao Tong university. He holds a master degree of thermal engineering.

Fleming Feng graduated from Tsinghua University at 1990, and then went to Shanghai Jiao Tong University for a master's degree. He started to work for Intel from year 1997 after 4 year experience in a software company. From 2000, Fleming started to work on Linux and open source project development. Ander from year 2002, Fleming began to work with TSP(the predecessor of OTC), managing the development team for Carrier Grace Linux, and from 2005, took the role for managing the OTC Shanghai team. From 2009, Fleming worked as the PRC chief OSS scientist majorly focusing on Linux and open source collaboration with China government, industry and university.