VIRTUAL EYE- ASSISTING VISUALLY IMPAIRED

[1]DEEPAK KUMAR SINGH, [2]KHUSHNASEEB, [3]MONIKA SWAMI, [4]RAVINDRA BHATT, [5]SHARAYU LOKHANDE

[1,2,3,4,5]Department of Computers, Army Institute of Technology
E-mail: deepakkumar0511@gmail.com, khushnaseeb.rockfreak@gmail.com, monika.swami.ait@gmail.com, ravindra.bhatt29@gmail.com, slokhande@aitpune.edu.in

Abstract- For blind people who have no visual information, there are a lot of dangers in everyday-life environments. Independent travel is well known to present significant challenges for such individuals, thereby reducing quality of life and compromising safety. This project aims to develop an intelligent, high performance, affordable and easy to use computer vision based visual information assistance system for blind. Visual information will be captured via two miniature cameras mounted on head via a cap or sunglasses, while image processing and speech synthesis will be done by a portable computer. Our detection results are from a large data set of indoor images including doors, staircases, cupboards, walls, etc. so that the proposed method is robust and generic to the changes of scales, view points, and occlusions in different buildings. An area based stereo matching is performed over the transformed images to calculate dense disparity image. Low texture filter and left/right consistency check are carried out to remove the noises and to highlight the obstacles. The audio is conveyed to the blind user through stereo headphones. The audio includes navigation assistance commands such as presence of obstacle, clear path, object of interest, doors nearby, other humans surrounding etc.

Keywords- Clear path, Edge Detection, Image Processing, Obstacle detection, Staircase Detection, Stereo Vision.

I. INTRODUCTION

Most aspects of the dissemination of information to aid navigation and cues for active mobility are passed to human through the most complex sensory system, the vision system. This visual information forms the basis for most navigational tasks and so with impaired vision an individual is at a disadvantage because appropriate information about the environment is not available.

Few of the technologies that have traditionally used for assisting in navigation include the white cane, guide dogs, guide robots, etc. These traditional methods have limitations such as limited range as in case of the cane, guide dogs might not trained well enough or certain places are inaccessible to guide robots. These methods also lack the fundamental concept vision as for humans and involve much less human perception.

This technology may not replace the cane, but should complement it: alert the user to obstacles a few meters away and provide guidance for going to a specific location in town or in a shopping center. With technical advancements in the field of computer vision and computing in general, new Vision based electronic aids have been devised. These include sensor based devices such as using IR Sensors, Ultrasounds, Lasers, etc. Sensor based devices have proven to be costly and in-affordable for most of the people. Others include SVETA, Acoustic perception based systems, etc. In this paper, we propose a Computer Vision based system which is robust and affordable.

II. PREVIOUS WORK

Different approaches exist to help the visually impaired. One system for obstacle avoidance is based on a hemispherical ultrasound sensor array. It can detect obstacles in front and unimpeded directions are obtained via range values at consecutive times. Talking Points is an urban orientation system based on electronic tags with spoken (voice) messages. These tags can be attached to many landmarks like entrances of buildings, elevators, but also bus stops and buses. A push-button on a hand-held device is used to activate a tag, after which the spoken message is made audible via the device's small loudspeaker.

Guide Cane is a computerized travel aid for blind pedestrians. It consists of a long handle attached to a sensor unit on a small, lightweight and steerable device with two wheels. Ultrasonic sensors detect obstacles and steer the device around them.

SWAN or System for Wearable Audio Navigation is a wearable computer with a variety of location and orientation-tracking technologies, including GPS, inertial sensors, pedometer, RFID tags, RF sensors and a compass. Sophisticated sensor fusion is used to determine the best estimate of the user's actual location and orientation. Tyflos-Navigator is a system which consists of dark glasses with two cameras, a portable computer, microphone, earphones and a 2D vibration array. It captures stereo images and converts them into a 3D representation. The latter is used to generate vibration patterns on the user's chest, conveying distances of the user's head to obstacles in the vicinity. Similar initiatives exploited other sensor solutions, for example an IR-multisensor array with smart signal processing for obstacle avoidance and a multi-sonar system with vibro-tactile feedback. One system is devoted to blind persons in a wheelchair. Information of the area around the wheelchair is collected by means of cameras mounted rigidly to it. Hazards such as obstacles, drop-offs ahead of or alongside the chair, veering paths and curb cuts can be detected for finding a clear path and maintaining a straight course. All camera information can be combined with input from other sensors in order to alert the user by synthesized speech, audible tones and tactile cues.

From the overview presented above we can conclude that technologically there are many possibilities which can be exploited. Some are very sophisticated, but also very complex and likely too expensive for most blind persons who, in addition to having to deal with their handicap, must make both ends meet financially. Moreover, ergonomically most may prefer not to wear a helmet or to use other visually conspicuous devices which set them apart.

III. PROPOSED SYSTEM

A. Acquire Image Data

The system consist of a set of cameras mounted together to form a set of stereo cameras. The scene information is captures with help of these cameras using Open Computer Vision Library. The acquired video feed is then processed on frame by frame basis using various functions from the OpenCV library to extract the required image features. Algorithms used in the process are derived from various sources (relevant references have been provided).

B. Clear Path Detection: Edge Detection

1) Path Detection

Virtual eye project uses a stereo camera to capture the environment in front of the user, which is worn by the user. The detection of clear path is based on: (a) establishing a region of interest for finding path; (b) pre-processing of frames to determine contours; (c) find the region with minimum ratio of edges.

IV. REGION OF INTEREST FOR PATH DETECTION

The input frames have a fixed width and height (W x H). The region of interest (ROI) is the area from the bottom of the frame to the height H/3 as we need to find the path in the region closest to the user. The width is divided in 3 overlapping areas as in figure 1, on which the processing is done. Each of these areas is gives an estimated clear path and based on the overlapping of these estimated regions an accurate clear path is determined.

Figure 1: Original image with ROI (in green)

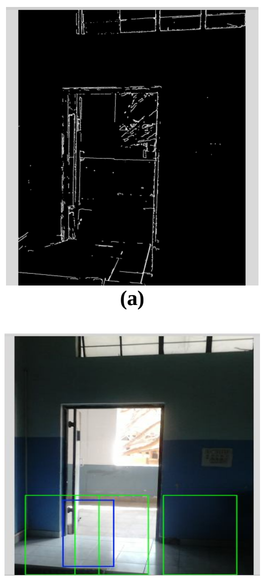
V. CONTOUR DETECTION

The Canny edge detector and contour detection are used to find the edges in the frames. In order reduce CPU processing time the processing is done on a grayscale version of the frames. The image is smoothened using a 3x3 normalized box filter.

The Canny edge detector is applied with Gaussian filter size as 3.0, in combination with TL = 161 and TH = 18 which are the low and high thresholds for hysteresis edge tracking. The result is a binary edge image. The image is then processed with Suzuki algorithm and TehChin algorithm to find the contours in the image.

2) Clear Path

The binary image containing the contours is then processed by dividing the ROI in 3 overlapping blocks as shown in figure 2. The blocks are processed to find the region with minimum ratio of white and black pixels. The result is three possible clear path in the ROI. The clear path is then determined as the largest region overlapping between any of the three regions. This is region is the path which the user has to take.



(a)

(b)

Figure 2: (a) Contours in image (b) Clear path (blue region)

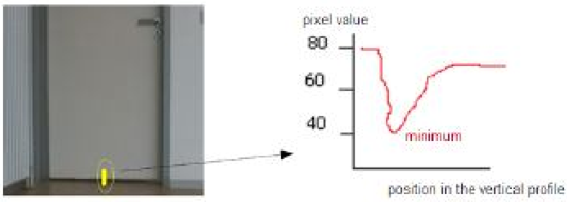
The path detection can be made more robust by using above approach along with the depth estimation using stereo camera. The objects like wall and other objects on close range may appear to have no edges and can be estimated as clear path but, this can be avoided by using a depth map of image.

C. Door Detection:

Doors are important landmarks for navigation because they provide the entrance and exit points of rooms, and because they provide stable and semantically meaningful structures for determining the location of a person. Most existing algorithms for door detection use range sensors or work in limited environments. There are two major problems in door detection. First it is often impossible to get the whole door in a single the camera image. Second door are diverse in nature, each door can have different color, pose, or lighting. Therefore a robust algorithm is required which addresses all the above mentioned issues. Robust vision-based door detection algorithms which utilize a variety of features including intensity edges. Geometric features like concavity and bottom-edge intensity profile increase performance significantly. We combined features from and to design a more robust and optimized door detector.

1) Algorithm
Input : Image
Output : Set of straight lines qualifying for a door

1. Extraction of Straight Lines: To group the edges into straight line segments, many existing approaches apply the Hough transform, but in our experiments we found this approach to be quite sensitive to the window size: Small windows cause unwanted splitting of lines, while large windows cause unwanted merging of lines. We used LSD to extract straight lines.
2. Almost vertical and horizontal lines i.e., having permissible deviations with y-axis and x-axis are kept. Then for each line segment we merge it with another line segment if the maximum tolerated angle deviation between them and the maximum distance between their end points are within a limit
3. Vertical lines which are parallel and are having length between certain thresholds qualify for height of the door, rest are discarded. Similarly we check for the width of the door.
4. Door Gap Test: Doors are constructed with a gap below them to avoid unnecessary friction with the floor. As a result, when the light in the room is on, the area under the door tends to be brighter than the immediate surroundings, whereas if the light is off, then the area tends to be darker. In either case this piece of evidence provides a surprisingly vital cue to the presence of the door.
For each pixel along the bottom door edge, we compute the minimum and maximum intensities in the surrounding vertical profile. If one of these extremes is above a threshold and located near the door edge, then the pixel votes for the presence of a door. A majority vote among the pixels is taken.



(a) (b)

Figure 3: (a) An image of a door. (b) The intensity profile of a vertical slice around the bottom edge. The dark region caused by the shadow of the door indicates the presence of the door.

5. Concavity Test: In many environments doors are receded into the wall, creating a concave shape for the doorway. A simplified concave door structure is illustrated below leading to two observations regarding the intensity edges:
• A slim "U" exists consisting of two vertical lines and one horizontal line.
• The bottom edge of the door is slightly recessed from the wall/floor edge.

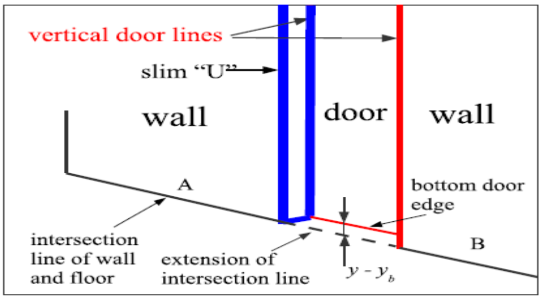


Figure. 4: A concave door exhibits a slim "U" to its side, as well as a recession of the bottom edge.

Let (x, y) be a pixel on the line formed by extending the wall/floor boundary in front of the door and let (x, y_{bottom}) be the pixel on the bottom of the door in the same column of the image. We test the recession of the door as follows:

$rec(\omega) = (\tau_{min} < y - y_{bottom} < \tau_{max})$, where $\tau_{min} = 2$ and $\tau_{max} = 10$ pixels, and $\omega \in \{L,R\}$ indicates whether the wall/floor boundary was extended from the left or the right of the door. The recession test is applied with respect to all non-vertical line segments below the vanishing point, and the slim "U" test is applied to the side of the edge. A door is declared if at least two of the three tests succeed

6. Check for the presence of Knobs: The distance between the vertical lines pairs is used to find the door knob area. In these two areas (left and right side of a door) presence of at least two almost horizontal lines indicates the presence of a knob.
7. All the above results are combined to indicate the presence of a physical door.

D. Staircase Detection

Stairs are one of the few obstacles that need to be detected for aid in navigation for blind persons. We aim to detect stairs both indoor and outdoor at small range, about 5m at maximum, and in front of the user. There are many factors that make the detection difficult like illumination, different textures, view angles and design. However the proposed method does detect a fair number of stairs leaving behind a few false positives.

1) Method

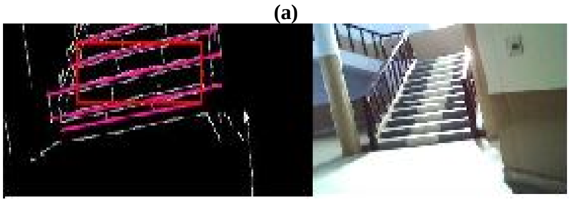
The main feature of stairs that we exploit for detection is almost parallel edges and the distance between the steps that varies almost linearly. These edges are almost but not necessarily horizontal. Since the system has to work in real time, performance needs are a big issue.

To optimize performance processing is done in a window or region-of-interest (ROI) which captures information in front of the camera, in the centre and at the bottom part, covering the immediate vicinity until a distance of approximately 5m and the captured frame is converted into a greyscale image.

Further the captured frame is processed as per method described in [9] with slight modifications.

Certain false positives cannot be avoided as of now and occlusions in the view of the staircase might cause false negatives.

(a)



(b)

Figure 5: Few Test Result of staircase detection

E. Object Recognition and detection

The object detector described below has been initially proposed by Paul Viola and improved by Rainer Lienhart. First, a classifier (namely a cascade of boosted classifiers working with haar-like features) is trained with a few hundred sample views of a particular object (i.e., a face or a car), called positive examples, that are scaled to the same size (say, 20x20), and negative examples - arbitrary images of the same size.

After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a "1" if the region is likely to show the object (i.e., face/car), and "0" otherwise. To search for the object in the whole image one can move the search window across the image and check every location using the classifier.

Figure 6: Detection results with haar-classifier

The classifier is designed so that it can be easily "resized" in order to be able to find the objects of interest at different sizes, which is more efficient than resizing the image itself. So, to find an object of an unknown size in the image the scan procedure should be done several times at different scales. The word "cascade" in the classifier name means that the resultant classifier consists of several simpler classifiers (stages) that are applied subsequently to a region of interest until at some stage the candidate is rejected or all the stages are passed. The word "boosted" means that the classifiers at every stage of the cascade are complex themselves and they are built out of basic classifiers using one of four different boosting techniques (weighted voting). Currently Discrete Adaboost, Real Adaboost, Gentle Adaboost and Logitboost are supported with OpenCV.

F. Disparity Map generation and Estimating distance to Obstacle:

1) Stereo Vision

Computer stereo vision is the extraction of 3D information from digital images, such as obtained by a digital camera. By comparing information about a scene from two vantage points, 3D information can be extracted by examination of the relative positions of objects in the two panels. This is similar to the biological process Stereopsis.

In traditional stereo vision, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene, in a manner similar to human binocular vision. By comparing these two images, the relative depth information can be obtained, in the form of disparities, which are inversely proportional to the differences in distance to the objects.

To compare the images, the two views must be superimposed in a stereoscopic device, the image from the right camera being shown to the observer's right eye and from the left one to the left eye.

In real camera systems however, several pre-processing steps are required.

1. The image must first be removed of distortions, such as barrel distortion to ensure that the observed image is purely projection.
2. The image must be projected back to a common plane to allow comparison of the image pairs, known as image rectification.
3. The displacement of relative features is measured to calculate a disparity map
4. Optionally, the disparity as observed by the common projection is converted back to the height map by inversion. Utilizing the correct proportionality constant, the height map can be calibrated to provide exact distances.

2) What is Disparity Map?

Disparity refers to the difference in location of an object in corresponding two (left and right) images as seen by the left and right eye which is created due to parallax (eyes' horizontal separation). The brain uses this disparity to calculate depth information from the two dimensional images.

In short, the disparity of a pixel is equal to the shift value that leads to minimum sum-of-squared-differences for that pixel. It is inversely proportional to depth, and it is possible to define a mapping from an (x, y, d) triple to a three-dimensional position.

3) Calculating Disparity Map

The images are padded with a "frame" of zero pixels to facilitate the window operation (SSD/SAD) at the border.

SSD - Sum of Squared Differences

SAD - Sum of Absolute Differences

First, squared difference or absolute difference is calculated for each pixel and then all the values are summed over a window W.

For each shift value of the right image, there is an SSD/SAD map equal to the size of the image. The disparity map is a 2D map reduced from 3D space. The disparity of a pixel is equal to the shift value that leads to minimum SSD/SAD for that pixel.

4) Calculating Disparity Map using OpenCv

In OpenCv, to calculate disparity map, it has StereoBM and StereoSGBM. StereoBM stands for block matching algorithm. Stereo SGBM stands for semi block matching algorithm. OpenCv implements Stereo Visions as described in references: .

G. Face Detection and Face Recognition

It is important to keep track of other humans in the user's environment and recognize them. Virtual Eye detects human faces as well as recognizes them. To this purpose we use Haar Classifiers, to rapidly detect any object, including human faces, using AdaBoost classifier cascades that are based on Haar-like features and eigenfaces for recognition.

The process for detecting and recognizing faces consists of three steps: (a) detecting faces in frame using haar cascades and extracting them; (b) using Fisher face recognizer to train the face database; (c) predict faces from the trained data.

1) Haar Cascade Classifiers

The core basis for Haar classifier object detection is the Haar-like features. These features use the change in contrast values between adjacent rectangular groups of pixels.

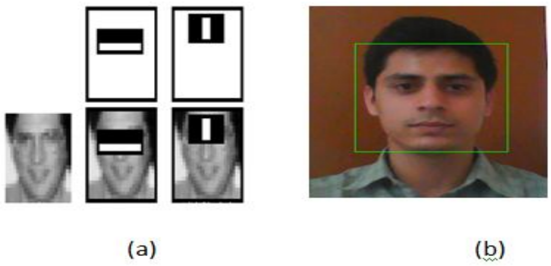


Figure 7: (a) Contrasting regions around the eyes

(b)Detected face

The contrast variances between the pixel groups are used to determine relative light and dark areas. For example the region around the eyes is darker than the regions above and below (Figure 7). Such features are used as distinct measures to detect faces in frames.

2) Integral Image

The simple rectangular features of an image are calculated using an intermediate representation of an image, called the integral image. The integral image is an array containing the sums of the pixels' intensity values located directly to the left of a pixel and directly above the pixel at location (x, y) inclusive. So if A[x, y] is the original image and AI[x, y] is the integral image then the integral image is computed as shown in equation 1 and illustrated in Figure 8.

$$AI[X,Y] = \sum_{x' \le x, yi \le y} A(x', y') \quad (1)$$

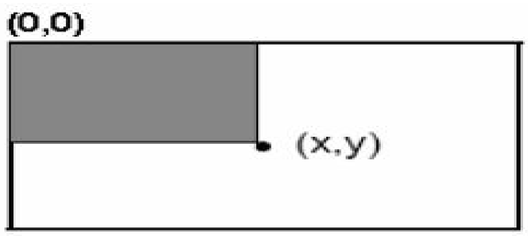


Figure 8: Summed area of integral image

3) Using Face recognizer To Train Database

Fisherfaces uses Linear Discriminant Analysis to performs class specific dimensionality reduction, the Linear Discriminant Analysis maximizes the ratio of between-classes to within-classes scatter, instead of maximizing the overall scatter. The idea is simple: same classes should cluster tightly together, while different classes are as far away as possible from each other in the lower-dimensional representation.

4) Recognizing faces in frames

The faces detected in the frame are then predicted using the trained face recognizer model. This returns the label of the subject if the face is recognized as in Figure 9. The Fisherface method handles variation in lighting and expression very well.

Figure 9: The recognized face

H. People detector:

One of the obstacles is human in the user's environment. Virtual Eye is capable of full body human detection. Although humans have been proven to be a much more difficult obstacle to detect because of their wide variability in appearance due to clothing, articulation and illumination conditions that are common in outdoor scenes. For this purpose we use default HOG cascade Classifiers available in OpenCV library. There are number of trained classifiers for detecting objects of a particular type, e.g. faces (frontal, profile), pedestrians etc. are also available. In the classifier HOG features is added to cascade detection algorithm.

CONCLUSION AND FUTURE SCOPE

In this paper, we present an assistant system named Virtual Eye. The system exploits OpenCv and audio prompts to help visually impaired in perceiving and locating objects in the nearby surrounding. In future we can add integrated maps (GPS Navigation) for better localization. Better processing units can be used for faster result. Also machine learning algorithm and neural network can be applied for the purpose of self-learning.

REFERENCES

[1] Byeong-Seok Shin, Cheol-Su Lim "Obstacle detection and avoidance system for visually impaired people" In Proc. 2nd Int. Workshop on Haptic and Audio Interaction Design, Springer, LNCS 4813, pp. 78-85, 2007.

[2] Jason Stewart, Sara Bauman, Michelle Escobar, Jakob Hilden, Kumud Bihani, Mark W. Newman "Accessible contextual information for urban orientation" In Proc. 10th Int. Conf. on Ubiquitous Computing, pp. 332-335, 2008.

[3] I. Ulrich, and J. Borenstein, "The guide cane-Applying mobile robot technologies to assist the visually impaired," IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 31, no. 2, pp. 131-136, 2001.

[4] Jeff Wilson, Bruce Walker, Jeffrey Lindsay, Craig Cambias, Frank Dellaert "SWAN: System for wearable audio navigation" In Proc. 11th IEEE Int. Symp. on Wearable Computers, pp. 91-98, 2007.

[5] Nikolaos Bourbakis, Despina Kavraki "A 2D vibration array for sensing dynamic changes and 3D space for blinds' navigation" In Proc. 5th IEEE Symp. on Bioinformatics and Bioengineering, pp. 222-226, 2005.

[6] Bruno Andò, Salvatore Graziani "Multisensor strategies to assist blind people: a clear-path indicator" IEEE Trans. on Instrumentation and Measurement, Vol. 58, No. 8, pp. 2488-2494, 2009.

[7] Sylvain Cardin, Daniel Thalmann, Frédéric Vexo "Wearable system for mobility improvement of visually impaired people" Visual Computer Journal, Vol. 23, No. 2, pp. 109-118, 2006.

[8] James Coughlan, Roberto Manduchi, Huiying Shen "Computer vision-based terrain sensors for blind wheelchair users" Computers Helping People with Special Needs, Springer, LNCS 4061, pp. 1294-1297, 2006.

[9] Detection of Indoor and Outdoor Stairs Somayeh Shahrabadi, Joao M.F. Rodrigues, and J.M. Hans du Buf Vision Laboratory, LARSyS, University of the Algarve, 8005-139 Faro, Portugal

[10] John F. Canny "A computational approach to edge detection" IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 679-698, 1986.

[11] Suzuki, S. and Abe, K., "Topological Structural Analysis of Digitized Binary Images by Border Following". CVGIP 30 1, pp. 32-46 (1985)

[12] Teh, C.H. and Chin, R.T., "On the Detection of Dominant Points on Digital Curve". PAMI 11 8, pp. 859-872 (1989)

[13] Zhichao Chen Stanley T. Birchfield "Visual Detection of Lintel-Occluded Doors from a Single Image" Electrical and Computer Engineering Department, Clemson University, Clemson, SC 29634.

[14] Jens Hensler, Michael Blaich, and Oliver Bittel "Real-time Door Detection based on AdaBoost learning algorithm" University of Applied Sciences Konstanz, Germany Laboratory for Mobile Robots Brauneggerstr. 55 D-78462 Konstanz.

[15] LSD: a Line Segment Detector" by Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall.

[16] Paul Viola, Michel Jones "Rapid Object Detection using Boosted Cascade of simple features" In IEEE Conference on Computer Vision and Pattern RecOgnition 2001.

[17] Emperical Analysis of Detection Cascade of Boosted Classifiers for Rapid Object Detection

[18] Birchfield, S. and Tomasi, C. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998.

[19] J.Y.Bouguet. MATLAB calibration tool. http://www.vision.caltech.edu/bouguetj/calib_doc.

[20] Hartley, R.I., Theory and Practice of Projective Rectification. IJCV 35 2, pp 115-127 (1999).

[21] Hirschmuller, H. Stereo Processing by Semiglobal Matching and Mutual Information, PAMI(30), No. 2, February 2008, pp. 328-341.

[22] Zhang. A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.

[23] Lienhart, R. and Maydt, J. An extended set of Haar-like features for rapid object detection. IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.

[24] Adolf, F. How-to build a cascade of boosted classifiers based on Haar-like features. http://robotik.inflomatik.info /other/ opencv /OpenCV_ObjectDetection_HowTo.pdf, June 20 2003

[25] Cristinacce, D. and Cootes, T. Facial feature detection using AdaBoost with shape constraints. British Machine Vision Conference, 2003.

[26] Viola, P. and Jones, M. Rapid object detection using boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[27] Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals Eugen. 7 (1936), 179–188.

[28] Navneet Dalal and Bill Triggs. "Histogram of oriented gradients for human detection". IEEE Computer Society Conference on Vol. 1 in 2005.

★★★