REAL-TIME EYE LOCALIZATION, BLINK DETECTION, AND GAZE ESTIMATION SYSTEM WITHOUT INFRARED ILLUMINATION

Bo-Chun Chen, Po-Chen Wu, and Shao-Yi Chien

Media IC and System Lab
Graduate Institute of Electronics Engineering and Department of Electrical Engineering
National Taiwan University, Taipei, Taiwan
Email: {coslate,pcwu,sychien}@media.ee.ntu.edu.tw

ABSTRACT

Gaze tracking systems have high potential to be used as natural user interface devices; however, the mainstream systems are designed with infrared illumination, which may be harmful for human eyes. In this paper, a real-time eye localization, blink detection, and gaze estimation system is proposed without infrared illumination. To deal with various lighting conditions and reflections on the iris, the proposed system is based on a continuously updated color model for robust iris detection. Moreover, the proposed algorithm employs both the simplified and the original eye images to achieve the balance between robustness and accuracy. Experimental results show that the proposed system can achieve the accuracy of 96.8% for blink detection and the accuracy of 1.973 degree for gaze estimation with the processing speed of 10–11fps. The performance is comparable to previous works with infrared illumination.

Index Terms**—** Eye tracking, eye localization, blink detection, gaze estimation

1. INTRODUCTION

Eye tracking/gaze tracking is a highly potential technique for natural user interface, especially for wearable devices. Most existing wearable eye trackers employ infrared illumination to achieve robust performance by stable lighting conditions with infrared light sources close to the eyes. Since infrared light is invisible, it does not distract users, and the pupil becomes obvious and easy to detect under infrared illumination. However, infrared light sources illuminating eyes in such a close distance may cause harms to eyes. Radiation in the near infrared light is the most hazardous and is transmitted by the optical components of eyes [1]. It has higher transmission rate and lower absorption rate than visible light, and near infrared light can eventually reaches the retina while visible light is always absorbed by cornea [2]. If one is overexposed to infrared light, it may cause the thermal retina burn [3].

In this work, we propose a new solution for eye tracking systems without infrared illumination. To deal with the variation of uncontrolled lighting conditions and maintain the accuracy as well, several strategies are proposed. First, an iris color model is utilized for robust iris detection, which is continuously updated to address various lighting conditions. Second, the proposed system employs a simplified and original eye images at the same time as a kind of multi-scale algorithm. The simplified image is employed to detect coarse seed point and mask impossible feature points. Third, a modified Starburst algorithm is proposed with an iris mask for eye images without infrared illumination.

This paper is organized as follows. Section 2 first introduces and reviews some previous related works. Next, Section 3 describes the proposed algorithm in detail. Section 4 then verifies the proposed algorithm and shows the experimental results. Finally, Section 5 draws the conclusions of this work.

2. PREVIOUS WORKS AND BACKGROUND KNOWLEDGE

Morimoto and Mimica [4] use infrared light sources and camera optical center to calculate the world coordinates of pupil and cornea. Hennessey [5] also uses infrared light sources to form reflection points, in other words, glints, which are used with the help of auxiliary coordinate of some already-known coordinates of object. The accuracy is approximately 1.5 degree with free-head movements. Sugano *et al.* [6] use saliency map and Gaussian Process Regression to realize auto-calibration. It is also a work without infrared illumination. Their accuracy is in the range of 1–3 degree through low-resolution input eye images. Lu *et al.* [7] use local binary pattern to train a regression with gaze points. Li *et al.* [8] simply use the eye center itself in a 2D image as features, and an effective algorithm called Starburst is proposed. Zhu *et al.* [9] use pupil/corneal reflection technique to get pupil-glint vector and then plus with 3D eye position, which is encoded in head movements information, as a feature to train a support vector regression as their mapping function. The accuracy of their work is in the range of 1–5 degree. In summary, most of the state-of-the-art works [10] [11] exploit extra infrared light sources, and the achieved accuracy is around 3 degree.

Blink of a person can be decomposed into two categories: voluntary blink and involuntary blink. Svensson *et al.* [12] detect involuntary blink as a way of drowsiness detection. Krolak and Strumillo [13] use an active model to get the eye region. They determine if the blink is voluntary by calculating the area of visible part of the eye, which is the region surrounded by the active curves. The overall accuracy is about 96%. Chau and Betke [14] use template matching for voluntary blink detection, where the matching function is the normalized correlation coefficient. The overall accuracy is 95.3%.

3. PROPOSED ALGORITHM

Fig. 1 shows the overview of the whole system. The input of the system is an eye image. Different from the eye images acquired under infrared illumination, the eye images under normal lighting conditions usually contain noise from lighting condition variation and reflection of environment, as shown in Fig. 2(a). Therefore, the first step is image pre-processing, which aims to filter out the distortion

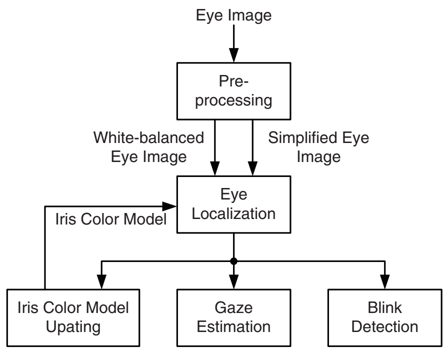
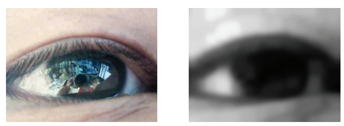


Fig. 1. System overview



(a) (b)

Fig. 2. (a) White-balanced eye image I_E. (b) Simplified eye image $\hat{I_E}$.

and noise of input eye image. Since the color information is critical to the whole system, we correct the color with automatic white balance to generate image I_E. Moreover, a simplified gray-scale eye image $\hat{I_E}$ is also generated with histogram equalization, morphological gray opening operation, and Gaussian blurring operation to reduce the reflection inside the iris, as shown in Fig. 2(b). As mentioned in Section 1, both these two images are employed in the eye localization as a kind multi-scale processing: the simplified image is used to generate rough detection and masking information robustly, while the white-balanced image is utilized to generated accurate eye location accordingly. An iris color modeled is also maintained in this system to address various lighting conditions. With the results and partial results of eye localization, the system can then estimate gaze and detect blink events. The details of the algorithm are described in the following subsections.

3.1. Eye Localization

The purpose of eye localization is to find a stable feature that can represent the location of an eye with high intra-class similarity and low inter-class similarity. We define the stable feature as a "virtual center." The virtual center in our system is not the location of pupil but is the center of the region surrounded by the feature points at eyelids and left/right sides of limbus. The method of feature point detection is based on Starburst algorithm [8], which is originally invented to detect the boundary of one's pupil under infrared illumination. Because the pupil is invisible under visible light, it is proposed to change the target to limbus and eyelids.

In order to make Starburst algorithm robust enough to detect feature points under visible light and have the ability to adapt to variant lighting conditions, two distinct features are developed and added. The first added feature is the automatic seed point finding algorithm. In the original Starburst algorithm, the seed point is picked as the center point of a frame. In order to automatically find the seed point with illumination invariant property, a color model with 2D h-s (hue-saturation) histogram is constructed for the iris region. During initialization, the rough iris region is estimated by calculating the moment center of the binarized version of $\hat{I_E}$ followed by a region growing process. Within the rough iris region, the h-s histogram of I_E is calculated as the initial iris color model. After the iris color model is established, for an input eye image I_E, the back projection of the iris color model is derived by replacing the value of each pixel (i, j) with the corresponding normalized bin value in the h-s histogram. That is,

$$I_{BP}(i, j) = H(h_{i,j}, s_{i,j}), \tag{1}$$

where the higher value means that the corresponding position has a higher probability to be a part of the iris region, as shown in Fig. 3(b). The seed point is then generated with binarization and moment center calculation. In the following frames, in order to deal with various lighting conditions, the iris color is updated every specified number of frames. To do so, a rough iris mask M_{RI} is generated by finding the convex hull of the binarized $\hat{I_E}$, as shown in Fig. 3(c). A new h-s histogram of I_E is calculated inside the mask M_{RI}. The back projection I_{BP} is then generated as well as an index called iris ratio, which is defined as.

$$\text{Iris Ratio} = \frac{\eta}{\sigma}, \tag{2}$$

where η means the sum of probability of iris point inside the mask M_{RI}, and σ means the sum of probability of iris point outside the mask. Greater iris ratio comes a more representative model, and we discard those 2D h-s histogram models with low iris ratio.

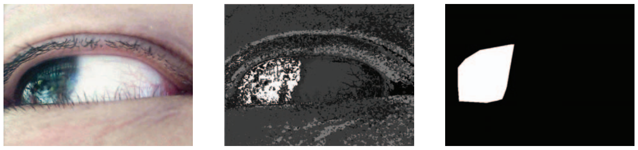
The second feature added is the iris mask that is used to mask out the reflection in iris region and other impossible feature points to enhance Starburst algorithm. As shown in Figs. 4(a)–(c), the iris region is estimated with projections of the simplified image $\hat{I_E}$. Eyelid fitting is then done by modeling the upper and lower eyelids as two parabolas. With the simplified image $\hat{I_E}$, the valley-peak field, Fig. 4(f), is generated by calculating the convex-hull of the biggest circle inside the extracted iris rectangle, as shown in Fig. 4(e), and the pixels with large horizontal gradient but low vertical gradient, as shown in Fig. 4(d). We then use the eyelid feature points outside the valley-peak field to fit upward and downward parabola functions, as shown in Fig. 4(g). The intersection of the region inside the eyelid and the estimated iris region forms the iris mask M_I for the modified Starburst algorithm described as Algorithm 1, and the feature points are then generated as shown in Fig. 4(h). Next, the accurate iris mask M_{AI} is generated based on the feature points, as shown in Fig. 4(i). Finally, the location of the virtual center (X_e, Y_e) is derived from the moment center of the inverted version of $\hat{I_E}$ masked by M_{AI}, as shown in Figs. 4(i) and (k) and the following equations.

$$m_{i,j} = \sum_{x,y} (255 - \hat{I_E}) \cdot M_{AI} \cdot x^i y^j, \tag{3}$$

$$X_e = m_{1,0}/m_{0,0}, Y_e = m_{0,1}/m_{0,0} \tag{4}$$

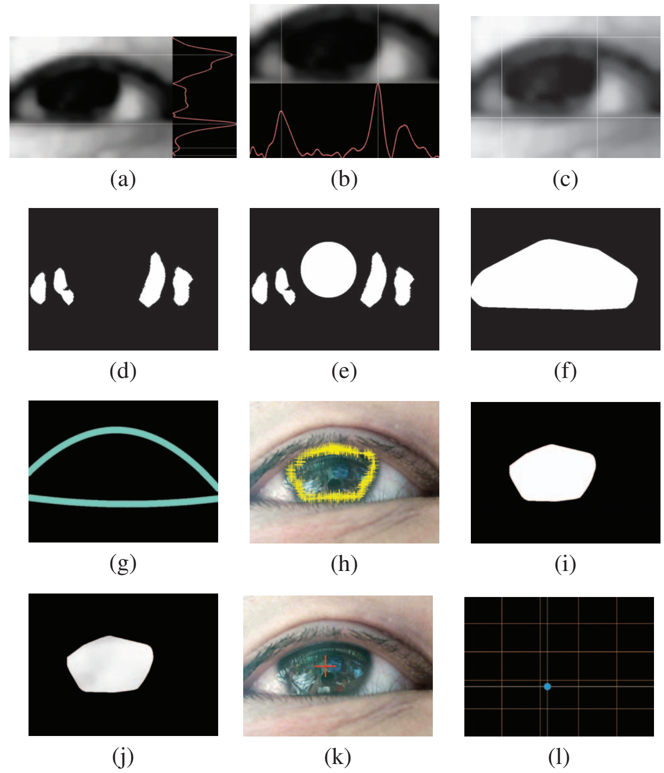
3.2. Gaze Estimation

The gaze estimation is mainly composed of two parts: calibration and multivariate polynomial interpolation. Calibration is the procedure to collect the mapping sets of virtual centers and gaze points. Several calibration points are displayed on the monitor, and the user is requested to look at each point for a while. Afterwards, several sets of correspondence in the form of $\{(X_e, Y_e), (X_s, Y_s)\}$ are generated, where the gaze point is marked as (X_s, Y_s) and the virtual center is marked as (X_e, Y_e). According to the position of the iris



(a) (b) (c)

Fig. 3. Hue-saturation histogram model refreshment (a) Input eye image. (b) Back projection I_{BP}. (c) Rough iris mask M_{RI}.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Fig. 4. (a) Input image and the horizontal projection. (b) Input image and the vertical projection. (c) Extracted iris rectangle. (d) High x-directed gradient but low y-directed gradient pixels. (e) Fig. 4(d) plus extracted biggest circle. (f) Valley-Peak field. (g) Fitted parabolas. (h) Detected feature points. (i) Accurate iris mask M_{AI}. (j) Inverted iris masked simplified eye image. (k) Virtual center. (l) Multivariate polynomial interpolated gaze point.

on the eyeball that is projected onto the image plane of the eye camera with radial-distorted lenses, the transformation from eye space to scene space is not linear. Therefore, we treat this mapping as a transformation from one curved surface to another, which is modeled as a multivariate polynomial function, where the order is empirically set as two in our system.

$$\overrightarrow{X_s} = \mathbf{M} \times \vec{a}, \overrightarrow{Y_s} = \mathbf{M} \times \vec{b}, \tag{5}$$

where

$$\mathbf{M} = \begin{pmatrix} 1 & X_{e1} & Y_{e1} & X_{e1}^2 & X_{e1}Y_{e1} & Y_{e1}^2 \\ 1 & X_{e2} & Y_{e2} & X_{e2}^2 & X_{e2}Y_{e2} & Y_{e2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{eL} & Y_{eL} & X_{eL}^2 & X_{eL}Y_{eL} & Y_{eL}^2 \end{pmatrix}, \tag{6}$$

$$\vec{a} = [a_0, a_1, a_2, \cdots, a_5]^T, \vec{b} = [b_0, b_1, b_2, \cdots, b_5]^T, \tag{7}$$

$$\overrightarrow{X_s} = [X_{s1}, X_{s2}, \cdots, X_{sL}]^T, \overrightarrow{Y_s} = [Y_{s1}, Y_{s2}, \cdots, Y_{sL}]^T. \tag{8}$$

By solving the model parameters $\vec{a}$ and $\vec{b}$, the gaze is estimated with the virtual center as the input, as shown in Fig. 4(l).

3.3. Blink Detection

The blink detection aims to detect one's voluntary blinking. With the information generated in the eye localization step, we simplify calculate the aspect ratio of the rectangle which bounding the mask M_{RI}. A one-dimensional median filter over frames is used to make the aspect ratio stable temporally. If the aspect ratio is too large, it means the eye closes because the detected object is the eyelids. If the aspect ratio is small and close to one, it means the eye opens because the shape of the iris is near a circle. We also take the appearance of limbus feature points as an input signal to determine whether the eye closes or opens. The limbus feature points are the feature points detected in Fig. 4(h) except for those inliers of upward and downward eyelid parabolas. As our eyes close, the limbus feature points will disappear. When the duration of a closing is larger than 5 seconds, we claim that one voluntary blink is detected; otherwise it is deemed an involuntary blink.

4. EXPERIMENTAL VALIDATION

Our system is implemented in Visual C++ and OpenCV library [15] on a personal computer with a 3.4-GHz CPU and 4-GB RAM. The prototype system setting is shown in Figs. 5(a) and (b), where the size of the monitor is 22 inches and the aspect ratio is 16:10($47.39 \times 29.61\ cm^2$). The resolution of an eye image is 640×480, and the resolution of the screen is 1680×1050. The visual angle ranges horizontally in $[-25.35, 25.35]$ $(degree)$ and vertically in $[-16.49, 16.49]$ $(degree)$. Tables 1 and 2 show the experimental results of the proposed blink detection algorithm. It shows that the proposed color model is effective, and the performance is comparable to the state-of-the-arts. Note that the comparison may be not fair because of different testing environments.

For the performance of gaze detection, the testing procedure is similar to that of calibration. The monitor shows several points, and the user is asked to gaze at each of these points for a while. The ground truth is the showed points, and the gaze error is defined as:

$$\text{Gaze Error (degree)} = \arctan\left(\frac{\Upsilon \times \psi}{\beta}\right), \tag{9}$$

where Υ is the error between the estimated gaze point and the ground truth in the unit of pixel, ψ is a constant that transforms the unit from a pixel to a centimeter, and β is the distance from the subjects to the monitor in the unit of centimeter. Fig. 5(c) shows one of the test results. Each of the 16 brown crosses with a red point at its center is the ground truth, and the green crosses around are the estimated gaze points while the subject gazes at the brown cross. Table 3 shows the experimental results with four subjects. Each of them goes through the test 5 times. The comparison with other works is shown in Table 4. It shows that the performance of the proposed system is similar to those of systems with infrared illumination. Again, the comparison may be not fair because of different testing environments. The processing speed of 10–11fps is achieved with the proposed system.

Input: Preprocessed eye image I_E and iris mask M_{AI}
Output: Feature points surrounding iris
initialization;
while *seed point does not converge* **do**
  clear F, the set of final feature points;
  Stage1:
  Emit rays radially from the seed point with angle ranging in $[0, 2\pi]$;
  for *Each ray* **do**
    Move extendedly along the ray from the seed point;
    Calculate derivation of intensity at each pixel;
    if *Outside the iris mask* $\wedge$ *derivation* > 0 **then**
      Push feature point to F;
    end
  end
  Stage2:
  for *Each candidate feature point detected in Stage 1* **do**
    Estimate the angle of the line from the feature point to the seed point, called Ang_{fs};
    Emit rays from the feature point back to the seed points with angle ranging in $[Ang_{fs} - \pi/12, Ang_{fs} + \pi/12]$;
    for *Each ray* **do**
      Move backward along the ray from the feature point;
      Calculate derivation of intensity at each pixel;
      if *Outside the iris mask* $\wedge$ *derivation* > 0 **then**
        Push feature point to F;
      end
    end
  end
  Seed point $\leftarrow$ Geometry center of feature points;
end
.

Algorithm 1: Modified Starburst algorithm.

5. CONCLUSION

In this paper, We propose a real-time solution for eye localization, gaze estimation, and blink detection without infrared illumination. We improve the seed point finding by using the 2D h-s histogram to address various lighting conditions. As the seed point becomes more robust, the whole Satrburst feature detection is also improved with the help of robust iris mask. The experiments conducted with the prototype system shows that the processing speed of the whole system is 10–11fps, the average gaze accuracy is about 1.973 degree over four subjects, and the overall accuracy of voluntary blink detection is 96.875%, which is comparable to those systems using multiple infrared light sources and multiple cameras. Note that, in this prototype system, the users have to fix their head positions during the tests. However, it can be further improved to a free-head-movement system by changing the display from the fixed screen to the display on the glasses or with calibration with the scene camera, which belongs to our future works.

6. ACKNOWLEDGMENTS

This work is supported by Ministry of Science and Technology, Taiwan (R.O.C.) under Grants MOST103-2221-E-002-268-MY3 and MOST103-2218-E-002-022.

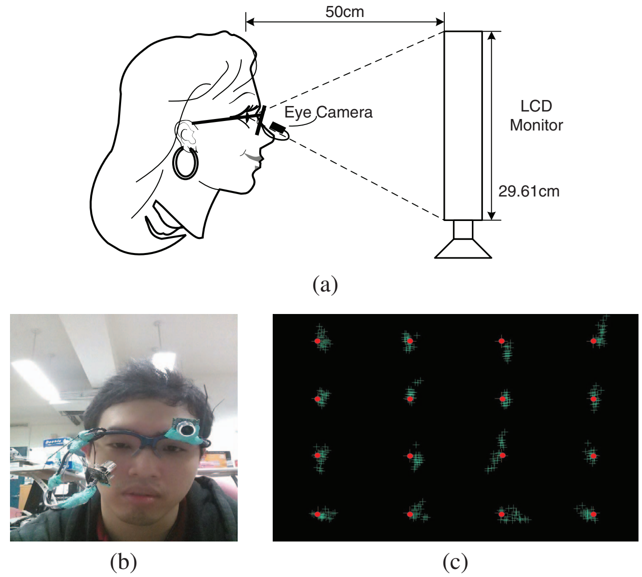


Fig. 5. (a) Environment setup. (b) Self-made glasses. (c) Test result of one subject.

Table 1. Test of voluntary blink detection for different subjects.

Item	With Color Model	Without Color Model
Subject 1	98% $\left(\frac{294}{300}\right)$	76.67% $\left(\frac{230}{300}\right)$
Subject 2	93.33% $\left(\frac{56}{60}\right)$	75% $\left(\frac{45}{60}\right)$
Subject 3	96.67% $\left(\frac{58}{60}\right)$	66.67% $\left(\frac{40}{60}\right)$
Subject 4	95% $\left(\frac{57}{60}\right)$	71.67% $\left(\frac{43}{60}\right)$

Table 2. Comparison with other works.

Item	Accuracy	Description
Krolak *et al.* [13]	96%	Active model
Chau *et al.* [14]	95.3%	Template matching
Proposed	96.875% $\left(\frac{465}{480}\right)$	

Table 3. Test of gaze estimation with four subjects.

Item	Mean	Standard Deviation
Subject 1	1.85641	0.196075
Subject 2	1.96989	0.22953
Subject 3	2.10671	0.461149
Subject 4	1.96165	0.125324

Table 4. Comparison with other works.

Item	Error(degree)	Description
Sugano et al. [6]	2 - 4	no IR
Noureddin et al. [10]	3	2 cameras, 1 IR
Guestrin et al. [11]	1 - 3	Single camera, 2 IR
Proposed	1.8 - 2	Single camera, no IR

7. REFERENCES

[1] F. Seeber, "Light sources and laser safety," *Fundamentals of Photonics*, 2007.

[2] Chern-Sheng Lin, "Discussion about harm that lasers and other souces of lights do to eyes," Oct. 1999, `http://140.134.32.129/scteach/scteach88/index.html/`.

[3] A.J. Allnutt, "Safety with lasers and other optical sources," *Optica Acta: International Journal of Optics*, vol. 28, no. 10, pp. 1312–1312, 1981.

[4] C. H. Morimoto and Marcio R.M. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 4 – 24, 2005, Special Issue on Eye Detection and Tracking.

[5] C. Hennessey, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of ETRA: Eye Tracking Research and Applications Symposium*. 2006, pp. 87–94, ACM Press.

[6] Y. Sugano, Y. Matsushita, and Y. Sato, "Appearance-based gaze estimation using visual saliency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 329–341, 2013.

[7] H.-C. Lu, G.-L. Fang, C. Wang, and Y.-W. Chen, "A novel method for gaze tracking by local pattern model and support vector regressor," *Signal Processing*, vol. 90, no. 4, pp. 1290–1299, Apr. 2010.

[8] D. Li, D. Winfield, and D. J. Parkhurst, "Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

[9] Z. Zhu, Q. Ji, and K.P. Bennett, "Nonlinear eye gaze mapping function estimation via support vector regression," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR 2006)*, 2006, vol. 1, pp. 1132–1135.

[10] B. Noureddin, P. D. Lawrence, and C.F. Man, "A non-contact device for tracking gaze in a human computer interface," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 52–82, 2005.

[11] E. D. Guestrin and E. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1124–1133, 2006.

[12] U. Svensson and Statens vtäg-och transportforskningsinstitut (Sweden), *Blink Behaviour Based Drowsiness Detection: Method Development and Validation*, VTI rapport. Swedish National Road and Transport Research Institute, 2004.

[13] A. Krolak and P. Strumillo, "Vision-based eye blink monitoring system for human-computer interfacing," in *Proceedings of 2008 Conference on Human System Interactions*, May 2008, pp. 994–998.

[14] M. Chau and M. Betke, "Real time eye tracking and blink detection with usb cameras," Tech. Rep., Boston University Computer Science Department, 2005.

[15] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.