Hand Gesture Recognition for Sign Language using Combination of Appearance-Based Features and SVMs

Marjan Shahchera
Department of Communication Engineering
East Azarbaijan Science and Research Branch, Islamic Azad University
Tabriz, Iran
Email: m_shahchera@yahoo.com

Hadi Seyedarabi
Faculty of Electrical and Computer Engineering
University of Tabriz
Tabriz, Iran
Email: seyedarabi@tabrizu.ac.ir

Mousa Shamsi
Faculty of Electrical Engineering
Sahand University of Technology
Tabriz, Iran
Email: shamsi@sut.ac.ir

Mohammad Reza Asharif
Faculty of Information Engineering University of the Ryukyus
Okinawa, Japan
Email: asharif@ie.u-ryukyu.ac.jp

***Abstract*— Hand gestures are powerful way for human communications. The proposed method is effectively combined of the following steps to detect hand gesture. The two methods of extracting Haar-like features and histogram of oriented gradient features (HOG) are applied and combination of those features forms a special feature vector. By adding advance half size Haar-like features to the basic Haar-like features and homomorphic filtering performance of Haar-like features improved. Also by applying the new Tan and Trigger preprocessing before HOG, sensitivity of lightening conditions have been reduced. Finally, linear multi-class support vector machine classification is used. The system is tested on Massey university American sign language (ASL) numeric and alphabetic hand gesture datasets, and system have been successfully able to recognize hand gestures with the recognition rate of 98% on numeric ASL and the recognition rate of 93% on alphabetic ASL. In addition recognition rate was 96% on National University of Singapore (NUS) hand gesture dataset with crowed backgrounds and variant lightening conditions.**

Keywords-Hand Gesture Recognition; American Sign Language (ASL); Haar-like Features; Histogram of Oriented Gradients (HOG); Support Vector Machine (SVM)

I. INTRODUCTION

Hand gesture recognition (HGR) is one of the important research areas in engineering and bioinformatics. The hand gesture recognition technology refers to using computer technology to analyze the hand gesture and extract effective and useful information. Hand gesture recognition can be useful from game control to robot control and from virtual environments to Smart home systems. Hand gesture recognition can be first step for easy, natural and cheap way for communicating human and computer systems without any device attached. Early techniques of hand gesture recognition was glove based methods that data was sending by special sensors of electronic gloves[1]. Then this gloves replaced by optical markers [2]. After that Vision Based methods presented that involved only a camera. These systems seems likely to complement biological vision by describing artificial vision systems. The problem of these systems is crowded background, lightening insensitive, person and camera independent to achieve best performance. Also, such systems must be optimized to gather the requirements, with accuracy and robustness. Appearance based approaches use image features to model the visual appearance of the hand. A simple approach that often used is to look for skin colored regions in the image[3]. Although skin color detection is very sensitive to lightening conditions and other skin like objects. The popular image features which are used to detect human hands and recognize gestures, include hand colors and shapes[4], local hand features [5]and optical flow [6]. Lowe [7] proposed an algorithm named Scale Invariant Feature Transform (SIFT) to extract individual invariant features from images that can be used to achieve dependable matching between different observations of an object. Recently, interest in approaches working with local invariant features and Haar like features are becoming more popular [8], [9], [10], [11].

II. SYSTEM OVERVIEW

The problem discussed in this paper is a vision based identification of the static hand gestures. The system deals with images of bare hands, which allows the user to interact with the system in a natural way and in no matter what environment.

A. preprocessing

By applying preprocessing before the features extraction, sensitivity of the features to the luminance conditions or shadows will decreased. The homomorphic filtering was improved the haar like features and the new preprocessing improved the HOG features of the images.

1) homomorphic filtering

Homomorphic filtering (HOMO) [12]is a well known normalization method where separate the illumination and the reflectance components. First, the input image is transformed into the logarithm domain and then into the frequency domain. At this point, the high frequency components are emphasized and the low-frequency components are reduced. As a final step the image is transformed back into the spatial domain by applying the inverse Fourier transform. Finally, inverse

978-1-4799-5359-2/14/$31.00 ©2014 IEEE

exponential operation yields an enhanced image. Review the whole process is in Fig. 1.

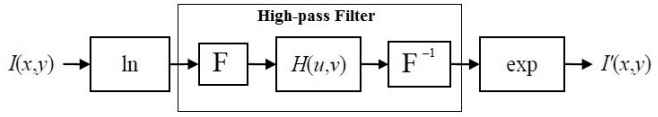
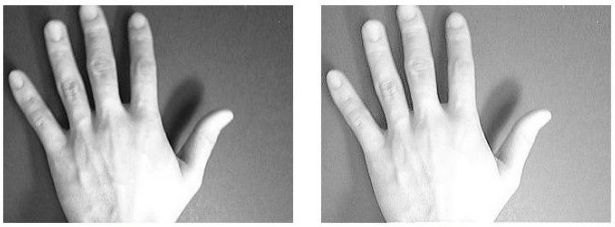


Figure 1. Homomorphic filtering block diagram

An example of the filtered image can be seen in Fig. 2. No filtering or histogram equalization is used in Fig. 2(a) and Fig. 2(b) shows homomorphic filtering results.



(a) (b)

Figure 2. (a) Original image, (b) Homomorfic filtering on image

2) Tan and Trigger preproessing

In this paper, we used the special preprocessing presented by Tan and Trigger [13]. Tan and Trigger indicate that the preprocessing can be successfully remove noise, shadow and illumination variation in human faces. This preprocessing have three steps. First step of preprocessing is gamma correction, it enhancing the local dynamic range of the image in dark or shadowed regions, while compressing it in highlights and bright regions. The basic principle is that the intensity of the light reflected from an object is the product of the incoming illumination L (which is piecewise smooth for the most part) and the local surface reflectance R (which carries detailed object-level appearance information). With gamma correction we want to recover object-level information independent of illumination[13].

Second step is Difference of Gaussian (DoG) Filtering. Gamma correction does not remove the influence of overall intensity gradients such as shading effects. Shading induced by surface structure is a useful visual cue but it is low frequency spatial information that is hard to separate from effects caused by illumination gradients. High pass filtering removes both the useful and the incidental information, thus simplifying the recognition problem and in many cases increasing the overall system performance. DoG filtering is a convenient way to obtain the resulting band pass behavior[13].

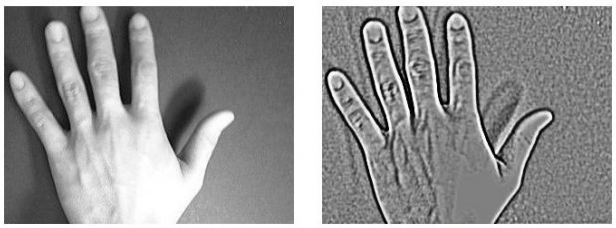
The final step is Contrast Equalization that globally rescales the image intensity to standardize a robust measure of overall contrast or intensity variation. It is important to use a robust estimator because the signal typically still contains a small admixture of extreme values produced by highlights, garbage at the image borders and small dark regions such as nostrils. One could, e.g., use the median of the absolute value of the signal for this, but here we have preferred a simple and rapid approximation based on a three stage process[13]:

$$I(x,y) \leftarrow \frac{I(x,y)}{(mean(|I(x',y')|))^{\frac{1}{\alpha}}} \quad (1)$$

$$I(x,y) \leftarrow \frac{I(x,y)}{(mean(\min(\tau,(|I(x',y')|)^{\alpha}))^{\frac{1}{\alpha}}} \quad (2)$$

$$I(x,y) \leftarrow \tau \tanh(\frac{I(x,y)}{\tau}) \qquad ,I \in (\tau,-\tau) \quad (3)$$

Here, α is a strongly compressive exponent that reduces the influence of large values, τ is a threshold used to truncate large values after the first phase of normalization, and the mean is over the whole (unmasked part of the) image. By default we use α = 0.1 and τ = 10. The resulting image is now well scaled but it can still contain extreme values. To reduce their influence on subsequent stages of processing, we finally apply a nonlinear function to compress over-large values. Here we use the hyperbolic tangent, thus limiting I to the range (τ,-τ)[13]. An example of the original and filtered image can be seen in Fig. 3.



(a) (b)

Figure 3. (a) Original image, (b) Tan & Trigger filtering on image

Example of one hand in different lightening condition and the corresponding Tan and Trigger preprocessing is shown in Fig. 4.

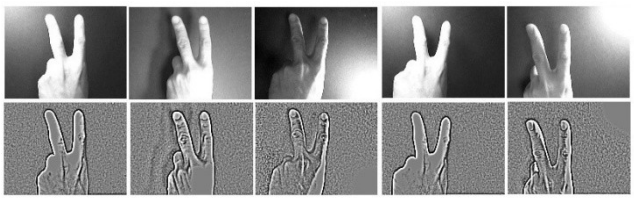


Figure 4. (upper row) Original images in different luminance conditions, (lower row) the corresponding Tan & Trigger filtering on images.

B. feature extraction

In this section we will explain features of image that we used for hand gesture detection and recognition. When the input data (such as a image) is too large to be processed then the input data will be transformed into a reduced representation set of features that also named features vector and transforming the input data into the set of features is called feature extraction. If the extracted features are carefully chosen it is expected that the features set will extract the useful information from the input data.

1) Haar like features:

Viola and Jones [14] used statistical method that can handle variations of human faces. The image features are rectangle features that called haar like features which reminiscent of haar basis functions. Each classifier is used unique rectangular areas to make decision if the region of the image looks like the model that trained or not. The value of a feature is the difference between the sums of the pixels within the black and white rectangle regions. Fig. 5(a) shows the basic haar like features that Viola and Jones presenting and

Fig. 5(b) shows the new additional set of haar like features that we presented in this paper.

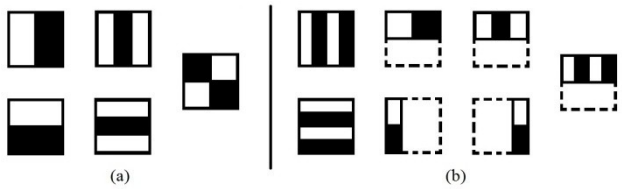


figure 5. (a) Basic haar like features, (b) new additional haar like features

In their method the concept of integral image is calculated using an intermediate representation of an image that accelerate the computation[14]. 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 p[x,y] is the original image and P[x,y] is the integral image then the integral image is computed as shown in (4) and illustrated in Fig. 6(a).

$$P(x,y) = \sum_{x' \le x, y' \le y} p(x', y') \quad (4)$$

According to the description of "Integral Image", the sum of the grey level value inside the area "D" in Fig. 6(b) can be computed as shown in (5).

$$P1+P4-P2-P3 = (A)+(A+B+C+D)-(A+B)-(A+C) = D \quad (5)$$

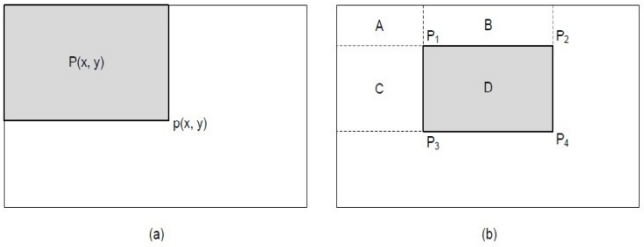


Figure 6.(a) The concept of integral image, (b) The sum of the pixels within rectangle D [14]

The basic haar like features are designed for face detection. face is symmetric and hand is asymmetric so for finding hand gestures, basic haar like features is useful. With adding these extra haar like features, finding finger positions in hand similar gestures is more possible.

2) HOG features:

The histogram of oriented gradients (HOG) illustrates the distribution of image gradients on different orientations and is implemented to capture shape and appearance feature in pedestrian, face and gesture detection [15]. The computation of HOG resemble of SIFT descriptor [7] and shape context. Each input image is divided into small equally sized regions called cells, then all pixels inside the cell vote in the histogram with its gradient magnitude as weights and the corresponding histogram reflects the pixel distribution with respect to gradient orientation, The same centered [-1, 0, 1] mask is used to compute horizontal gradient $p_x(x,y)$ and vertical gradient $p_y(x,y)$ of every pixel. Finally the histogram of each cell, which is presented as a vector, form the overall HOG feature for the image[8]. To reduce the illumination variance in different images, the new preprocessing is performed and it helped that all images have the same intensity range and no shadows.

C. classification

Support Vector Machines (SVMs) are a set of related supervised learning methods used for classification and regression[16]. Given a set of training examples, each belonging to one of two classes, a SVM training algorithm make a model that predicts whether a new example belongs into one class or the other. There are generally three ways to solve the problem: one against all, pair wise, and simultaneous classifications. In one against all classification, Vapnik proposed, to use continuous decision functions, instead of discrete decision functions. In pair wise classification, the *n*-class problem is changed into $n(n-1)/2$ two class problems[17]. In simultaneous classification we need to determine all the decision functions at once, which results in simultaneously solving a problem with larger number of variables than the above mentioned methods. Here we used pair wise separation method for classifying.

III. EXPERIMENTS AND RESULTS

We used a colored static hand gesture image dataset of numeral and alphabetical ASL gestures from Massey university as shown in Fig. 7 and Fig. 8.

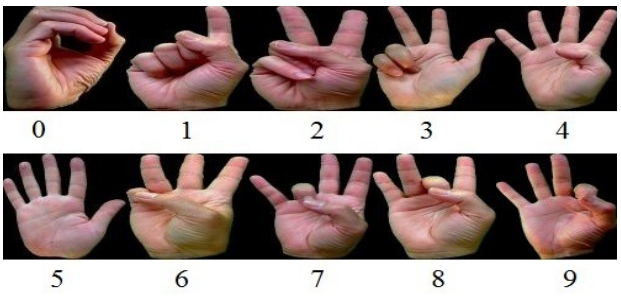


Figure 7. The complete ASL numeric set with sample segmented images

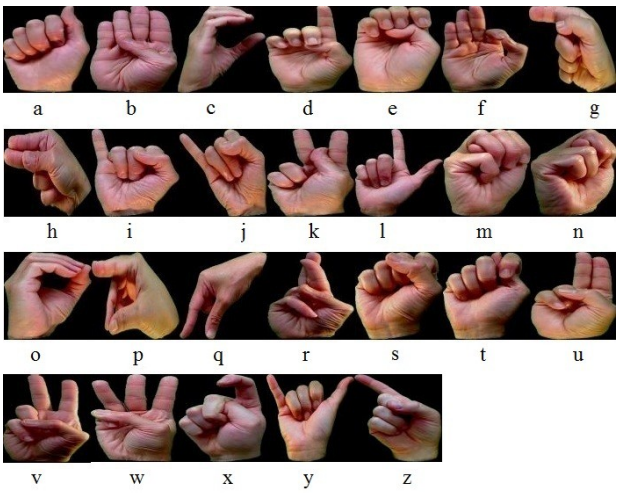


Figure 8. The complete ASL letter set with sample segmented images

We conducted our experiments with 55 samples of each numbers from 0-9 and 26 alphabets . The system is trained using 36 samples and test with 18 samples of each number and alphabet. As shown in Fig. 9 in numeric ASL dataset, basic haar like features (BH) can detect gestures with recognition rate of 54% and with adding homomorphic preprocessing (HOMO) recognition rate increased to 59%. By using new 12

haar like features (NH) that consist of 5 basic haar like features and 7 new haar like features the classification rate is 65% and with homomorphic preprocessing the classification rate increases to 77%. Therefore finding finger positions with new features in gestures that are so similar such as number (2,3,6,9) or number (4,7,8) is helpful. The HOG features with new preprocessing can detect hand gesture with rate of 90% and combination of these methods improved recognition rate to 98% .

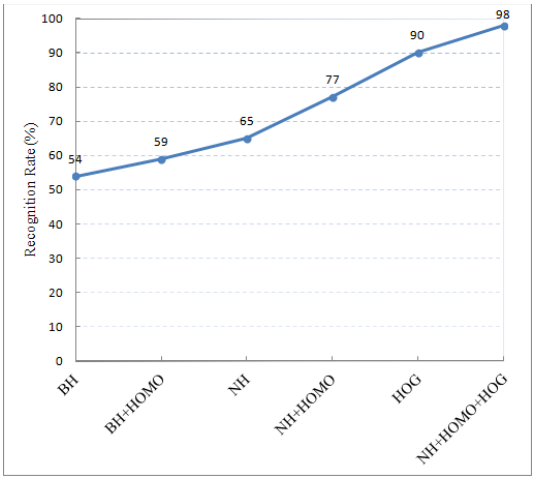


Figure 9. Overall results for the proposed methods on Massey university numeric ASL dataset, BH used for basic haar method, NH used for new haar method and HOMO used for homomorphic preprocessing.

As shown in Fig. 10 in alphabetic ASL dataset, basic haar like features (BH) poorly identify gestures with recognition rate of 23% and by using new 12 haar like features (NH) the recognition rate is 52% and with homomorphic preprocessing the classification rate increases to 60%.Some similar gestures such as (a, e, m, n, s, t) or (k, v) or (g, h, p) makes identifying so hard. The HOG features can detect hand gesture with rate of 85% and mixture of these methods enhanced recognition rate to 93%.

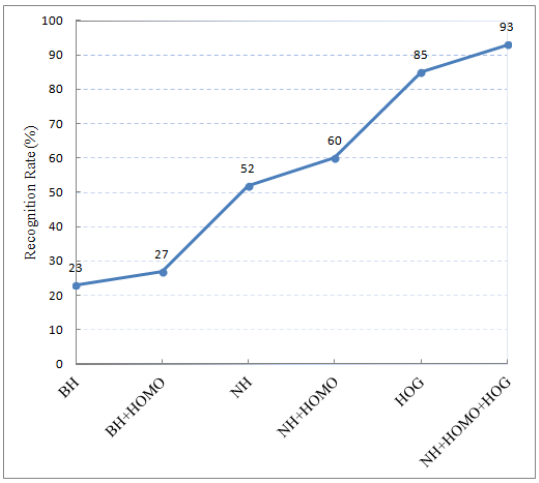


Figure 10. Overall results for the proposed methods on Massey university alphabetic ASL dataset

Furthermore for testing the proposed method in crowd background and difference illumination condition we use National University of Singapore (NUS) dataset with 4 example gestures as shown in fig. 11.

Figure 11. The samples of 4 gestures from NUS dataset with crowded backgrounds

In this dataset we use a simple skin detection method based on skin color detection in HSV color space [3] for finding hand region as shown in Fig. 12.

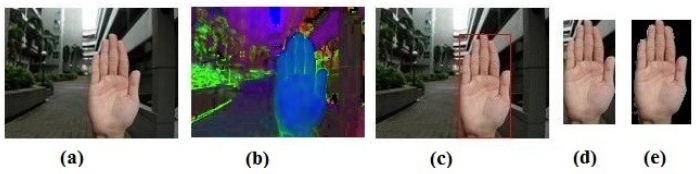


Figure 12. (a) Original image, (b) Image in HSV color space, (c) detected area of skin, (d) clipped skin area without background subtraction, (e) clipped skin area with background subtraction.

The system is trained using 60 samples and test with 40 samples in NUS dataset. Result of proposed methods with skin detection and without background subtraction is shown in Fig. 13 and result of proposed methods with skin detection and background subtraction shown in Fig. 14.

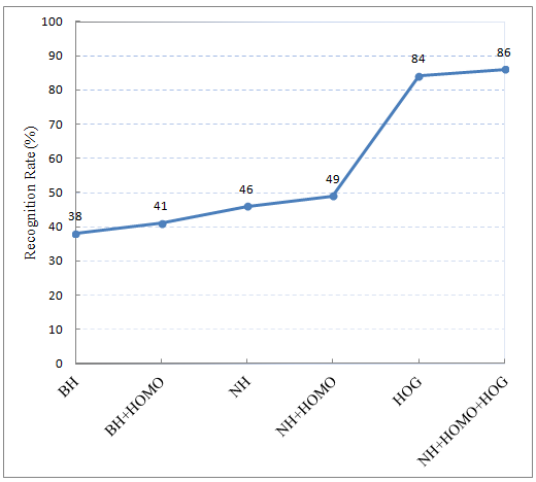


Figure 13. Overall results for the proposed methods on NUS dataset with skin detection and without background subtraction

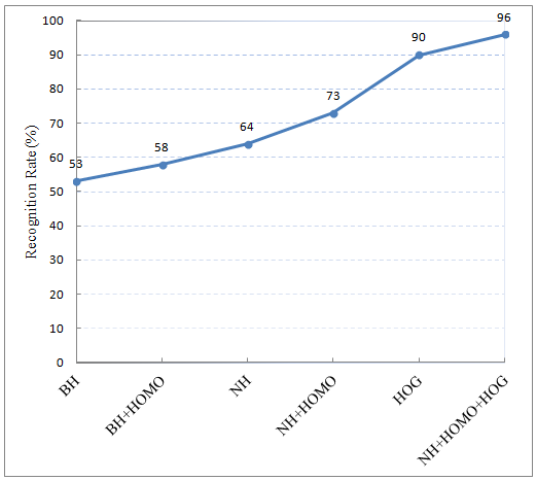


Figure 14. Overall results for the proposed methods on NUS dataset with skin detection and background subtraction

IV. CONCLUSION

We presented a new mixed method for hand gesture recognition under uncontrolled lighting based on robust preprocessing and combining powerful feature extraction methods. In our approach new half window size haar like features are presented for enhanced finger detection in similar hand gestures. Besides combination of haar like features and HOG features improved performance of system. We developed the results by applying homomorphic preprocessing and Tan and Trigger preprocessing by reducing sensitivity of system to lightening condition and shadows. The experimental results for combination methods in Massey university ASL numeric dataset have recognition rate of 98% and recognition rate in alphabetic letters was 93%. NUS dataset with and without background subtraction has recognition rate of 86**%** and 96%. In the next testing, another vision based method suggested to improve and extend the system, also using principle component analysis (PCA) to perform feature selection and decrease feature vector . Moreover testing system in real time human and computer environments is more advanced and interesting.

REFERENCES

[1] Sturman, D. J., & Zeltzer, D. (1994). “A survey of glove-based input”. Computer Graphics and Applications, IEEE, 14(1), 30-39.

[2] Chang, L. Y., Pollard, N. S., Mitchell, T. M., & Xing, E. P. (2007, October). Feature selection for grasp recognition from optical markers. In Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on (pp. 2944-2950). IEEE.

[3] Hasan, M. M., & Misra, P. K. (2011, June). Gesture recognition using modified HSV segmentation. In Communication Systems and Network Technologies (CSNT), 2011 International Conference on (pp. 328-332). IEEE.

[4] Jinda-apiraksa, A., Pongstiensak, W., & Kondo, T. (2010, May). A simple shape-based approach to hand gesture recognition. In Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on (pp. 851-855). IEEE.

[5] Bretzner, L., Laptev, I., & Lindeberg, T. (2002, May). Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. In Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on (pp. 423-428). IEEE.

[6] Simion, G., V. Gui, and M. Otesteanu. "Vision Based Hand Gesture Recognition: A Review." INTERNATIONAL JOURNAL OF CIRCUITS, SYSTEMS AND SIGNAL PROCESSING (2009).

[7] Lowe, David G. "Object recognition from local scale-invariant features." Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee, 1999.

[8] Wang, C. C., & Wang, K. C. (2008). Hand Posture recognition using Adaboost with SIFT for human robot interaction. In Recent progress in robotics: viable robotic service to human (pp. 317-329). Springer Berlin Heidelberg.

[9] Lienhart, R., & Maydt, J. (2002). An extended set of haar-like features for rapid object detection. In Image Processing. 2002. Proceedings. 2002 International Conference on (Vol. 1, pp. I-900). IEEE.

[10] Barczak, A. L., & Dadgostar, F. (2005). Real-time hand tracking using a set of cooperative classifiers based on Haar-like features.

[11] Chen, Q., Georganas, N. D., & Petriu, E. M. ,2008, Hand gesture recognition using Haar-like features and a stochastic context-free grammar. Instrumentation and Measurement, IEEE Transactions on, 57(8), 1562-1571.

[12] Delac, K., M. Grgic, and Tomislav Kos. "Sub-image homomorphic filtering technique for improving facial identification under difficult illumination conditions." International Conference on Systems, Signals and Image Processing. Vol. 1. 2006.

[13] Tan, X., & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lightening conditions. Image Processing, IEEE Transactions on, 19(6), 1635-1650.

[14] Viola, P., & Jones, M. (2001). “Rapid object detection using a boosted cascade of simple features”. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on (Vol. 1, pp. I-511). IEEE.

[15] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on (Vol. 1, pp. 886-893). IEEE.

[16] Abe, S. (2010). Support vector machines for pattern classification. Springer.

[17] Li, Z., Tang, S., & Yan, S. (2002). Multi-class SVM classifier based on pairwise coupling. In Pattern Recognition with Support Vector Machines (pp. 321-333). Springer Berlin Heidelberg.