World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:7, No:4, 2013

Exploiting Global Self Similarity for Head-Shoulder Detection

Lae-Jeong Park, and Jung-Ho Moon

Abstract**—**People detection from images has a variety of applications such as video surveillance and driver assistance system, but is still a challenging task and more difficult in crowded environments such as shopping malls in which occlusion of lower parts of human body often occurs. Lack of the full-body information requires more effective features than common features such as HOG. In this paper, new features are introduced that exploits global self-symmetry (GSS) characteristic in head-shoulder patterns. The features encode the similarity or difference of color histograms and oriented gradient histograms between two vertically symmetric blocks. The domain-specific features are rapid to compute from the integral images in Viola-Jones cascade-of-rejecters framework. The proposed features are evaluated with our own head-shoulder dataset that, in part, consists of a well-known INRIA pedestrian dataset. Experimental results show that the GSS features are effective in reduction of false alarmsmarginally and the gradient GSS features are preferred more often than the color GSS ones in the feature selection.

Keywords**—**Pedestrian detection, cascade of rejecters, feature extraction,self-symmetry, HOG.

I. Introduction

PEDESTRIAN detection from images has a variety of practical applications such as video surveillance and driver assistance system,so much research have been done actively in computer vision community and significant progresses have been made [1], [2]. However, pedestrian detection is still a challenging task because of complex backgrounds, variations of human poses, and clothes. It is even more difficult to detect pedestrian in crowded environments such as shopping malls and subway stations owing to frequent occlusion. In order to detect humans in those circumstances, there have been research on detection of upper parts of human body such as head and shoulders [3]-[7].

In particular, head-shoulder detection has attracted attentions as the head-shoulder pattern has more salient, informative shape than the head part only. In [4], a local shape matching scheme, which computes the resemblance with called edgelet features, has been proposed to detect and track multiple, partially occluded humans. In [5], [6], cascade-style head-shoulder detectors based on histogram of oriented gradients (HOG) features [8] rather than local silhouette matching, have been used to perform rapid human detection in crowded scenes. In [7], an aggregation of multilevel HOG and local binary pattern (LBP) features have been adopted to count the number of people in case where partial occlusion happens.

L.-J. Park and J-H Moon are with the Department of Electronic Engineering, Gangneung-Wonju National University, 120 Gangneung-Daehangno, Gangneung, Gangwon 210-702, South Korea (phone: 82-33-640-2389; e-mail: ljpark@gwnu.ac.kr).

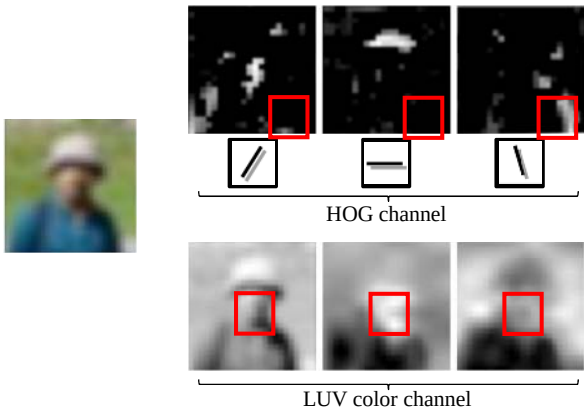


Fig. 1 Image channels for which local features are computed by some transformation such as sum and histogram

Meanwhile, it is also important for head-shoulder detector to operate in the near real-time, e.g., at a speed of 10-20 frames per second for practical applications. To accomplish this, cascade-of-rejecters approach, which is proposed by Viola and Jones (VJ) [9], has been widely used together with *local*, *rapid-to-compute* features such as HOG [10]-[11].In most VJ cascade classifiers, each classifier determines whether to pass or reject on the basis of not global features but local features. Because of such local features, the cascade classifiers sometimes would yield false alarms for input patterns that seem unlikely to be misclassified as head-shoulder patterns, as shown in Fig. 2. This is a limitation that the local features have inevitably, even if used in combination.

In this paper, we introduce a new class of features that exploits the global self-symmetry (GSS) characteristic in the head-shoulder pattern to reduce the false alarms effectively in local feature-based cascade classifiers. The features encode the similarity or difference of color and gradient information between blocks that are vertically symmetric to each other.

The remainder of this paper is organized as follows. Section II describes the global self-symmetry features in detail, and our cascade classifier is introduced briefly in Section III. Section IV discusses experimental results and conclusions are made in Section V.

II. Global Self-Symmetry Features

Fig. 1 shows two image channels, HOG and color for which the local features are extracted from a set of rectangular blocks by using integral images [9]. HOG is computed by building a weighted orientation-based histogram of gradients at pixels within a rectangular block. Likewise, a color histogram is built by summing color values within a block. In training a cascade

classifier, features from the feature pool, which consists of HOG and color histogram for every rectangular block, are evaluated and then the best ones are used as input features at each stage.

Even though the local features are known to be effective, they are not perfect enough to detect the head-shoulder patterns with few false positives. Fig. 2 shows examples of wrongly-detected negative images by a cascade classifier with HOG and color features. They share the same or similar shapes or edges locally with head and/or shoulder patterns, so it is difficult for the local features, chosen by the cascade classifier, to discriminate them correctly. However, they have clues to correct classification thanks to their rather *global* shapes that are quite different from the head-shoulder pattern. Most head-shoulder patterns have a global self-symmetry characteristic or strong vertical symmetry of shapes and colors around heads and both shoulder, even though slight variation of the symmetry occurs depending on human pose.

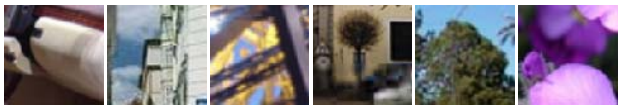


Fig. 2 Examples of false positive images

By exploiting the global, vertical symmetric characteristic, we propose global self-symmetric features, called GSS features. Fig. 3 illustrates the global, vertical symmetries of color and gradient information in the rectangular blocks around shoulders. It is probable that colors of clothes around both shoulders are vertically symmetric to each other. Besides, local shapes around both shoulders and/or head, which can be encoded by gradient histogram, are very likely to be vertically symmetric to each other. The shape self-similarity of the symmetric regions, e.g., the two blocks including the arrows in Fig. 3, can be extracted readily by re-using the HOG features.

Fig. 3 Global, vertical self-similarities of color and gradient information around shoulders

To be specific, the color GSS feature ismade by computing the affinity between local color histograms of the two symmetric blocks. Given a block, the corresponding block is selected randomly from a block pool that consists of blocks of various sizes within a region that is roughly vertically symmetric to the block specified. There are some well-known methods of measuring the similarity of two histograms, e.g., the L_1-norm, L_2-norm, and histogram intersection. Instead, we choose a similarity function that returns not a single value but a histogram whose element is given by

$$\frac{1}{\varepsilon + |p_i - q_i|}, \quad i = 1,2,\ldots,D, \tag{1}$$

where p_i and q_i denote the *i*-th elements of two histograms, D is the dimension of the histogram and ε is a control parameter. Preliminary experiments showed that the similarity function gave better performance than common histogram similarity measures. The shape GSS feature, called HOG GSS feature is computed from the two symmetric HOG blocks in the same way as the color GSS feature, except that the *i*-th element of a HOG block do *not* correspond tothe *i*-the element of the other block. Fig. 4 illustrates the one-to-one correspondence between orientation bin indices in two HOG blocks that are vertically symmetric to each other. For example, the HOG index 1 of a left-side block corresponds to the HOG index 6 of a right-side block. Based on the correspondence, the HOG GSS features are constructed by computing the vertical symmetry between the two HOG blocks.

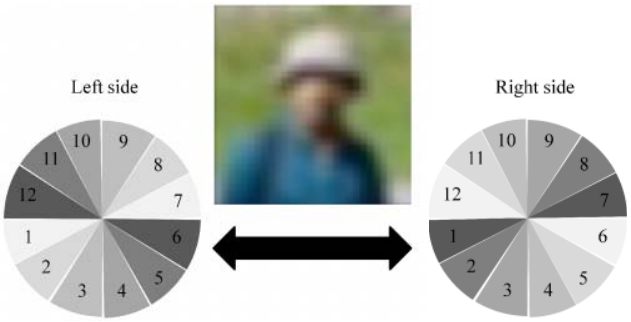


Fig. 4 The one-to-one correspondence of orientation binsin case of 12 orientation bins and 360 degrees

In [12], a self-similarity has been first incorporated into the feature extraction in pedestrian detection. The feature utilizes color similarities of body parts, e.g., arms and clothes by using color histogram intersections of all combinations of blocks. The vertical shape self-similarity has been utilized in [13] in HOG computation by using pixel weighting scheme based on the similarity. To our knowledge, it has not been attempted in pedestrian to incorporate the global, vertical self-symmetry into the HOG block level.

III. Cascade-Correlation Cascade Classifier

We choose the Viola-Jones (VJ) cascade-of-rejecters framework although the cascaded classifier has been known to be less competitive than support vector machines (SVMs).The merit of the cascaded classifier isthat it operates faster than SVMs as it uses simple classifiers which require much less computation load than SVMs. We adopt a modified cascaded classifier to address problems that VJ cascaded classifiers have in training: they fail to satisfy the performance criterion, e.g., false positive rate of 10^{-3}, and it takes a huge amount of training time in the order of several days. To enhance the training success rate and accelerate the training, we use *feed forward* connection through which the outputs of preceding classifiers are fed as inputs to the current stage classifier. The architecture

of the new classifier is illustrated in Fig. 5. The cascaded classifier consists of a series of classifiers, $H_i, i = 1,2,3...$, each determining whether to rejecter an input image. An input is rejected at the *i*-th stage if the output is larger than a threshold, and is passed into the (*i*+1)-th stage classifier otherwise. An input is classified as positive one if it has passed through classifiers at all stages.The output of the *i*-th stage classifier is represented by

$$H_i(\vec{x}) = \sum_{k=1}^{N_i} \alpha_k{}^i \cdot f_k{}^i + \sum_{k=1}^{i-1} \beta_k{}^i \cdot H_k(\vec{x}) \quad (2)$$

where $f_k{}^i$ and N_i denote the *k*-th feature and the number of features at the *i*-th stage, respectively. Feature $f_k{}^i$ is computed from a *D*-dimension vector, e.g., color histogram and gradient histogram in a feature block by using a *D*-to-1 mapping such as principle component analysis (PCA). Here, $f_k{}^i$ is computed by projecting a histogram into a discriminant vector determined by linear discriminant analysis (LDA). We choose the LDA-based *D*-to-1 mapping because it is very fast to compute in comparison with SVMs that has been used as weak classifiers in previous works, e.g., [10].The parameters, α and β are determined in a greedy style, unlike in boosting-style learning where they are by a formula according to error rates of weak classifiers.

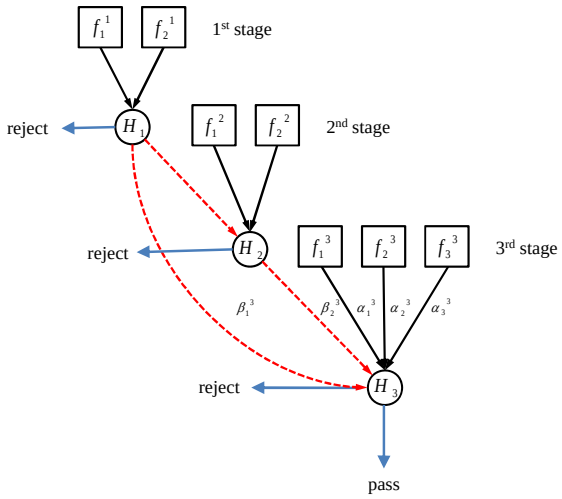


Fig. 5 The architecture of cascade-correlation classifier

It is quite reasonable that the outputs of preceding stage classifiers contribute to the training of the current stage classifier because, although training samples at the current stage (both positives and negatives) have passed through the preceding stage classifiers, they are likely to be distributed in the output space in favor of classification. Experimental results showed that the feedforward connections are effective in completing the training at much higher performance point as well as at a faster speed than without the connections. In fact, the concept of the feedforward connection between stages or layers originates from a neural network model, called cascade-correlation neural networks [14]. Hence, we call the monolithic cascaded classifier *cascade-correlation classifier*. A similar approach has been adopted in the VJ cascade classifier architecture, named nested cascade classifier [15] where each classifier receives the output of the last stage classifier only. The GSS features are evaluated in the cascade-correlation classifier framework.

IV. Experimental Results

A. Dataset

Because there is no open dataset for head-shoulder detection, we created a dataset by cropping pedestrian images from a public INRIA dataset [8] and adding about 120 negative images into the INRIA negative image set. While cropping head-shoulder patterns from the INRIA dataset, care must be made to center head patterns in the cropped images. Unlike pedestrian detection in case that full-body pattern is available, the head position is very a salient discrimination feature because the within-class variation of positive samples gets large if the head locations are aligned imprecisely.

The dataset consists of 1,800 positive samples with the size of 24x24 (1,000 samples for training and 800 for test) and 900 head-shoulder-free images (500 for training and 400 for test). The negative images of shopping mall, subway platform, and campus were added into the negative image set since the places are where we want to deploy the detector. In addition, images of lower-body parts such as legs and feet were included in the negative image set because preliminary experiments discovered those patterns often produce false alarms. Some positive and negative samples in the dataset are shown in Fig. 6.

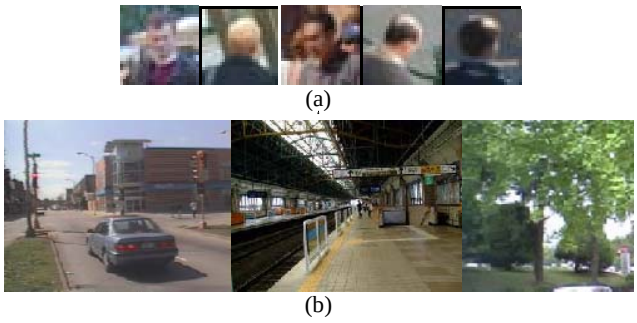


Fig. 6 A head-shoulder dataset. (a) positive samples, (b) images for negative samples

B. Results

A cascade-correlation classifier was trained in such a way that a classifier at each stage is trained by using 1,000 24x24-sized positive samples and 8,000 24x24-sized negative samples. Before used in training, the negative samples were selected randomly from the negative image set and passed through the preceding stage classifiers.

Table I shows the values of training parameters. The maximums (minimums) of height and width of feature blocks were set to 10 (3) since features from larger blocks were not chosen at all in the feature selection in training. For features,

TABLE I
TRAINING PARAMETERS

Parameters	Value
Image size	24 x 24
min/max of heights of feature blocks	3/10
min/max of widths of feature blocks	3/10
true positive rate at each stage	0.90
false positive rate at each stage	0.65
max no. of features at each stage	40
ε	1.0

signed HOG (with signed gradient information) features with 18 orientation bins were computed and CIE-LUV color histograms were also extracted for each block. All cascade-correlation classifiers, which trained independently with different sets of distinct features, were tested with 800 positive samples and 300,000 patches selected randomly from the test negative image set. The training took about 8-9 hours on Intel 2.7GHz CPU.

Performance was evaluated on graphs of miss rate vs. false positive per window(FPPW). Fig. 7 shows the performance of the classifiers with HOG features only (dotted line), HOG+LUV features (dashed line), and HOG+LUV+GSS features (solid line). For fair comparison, all curves were averaged for 10 classifiers, each trained independently. Surprisingly, adding LUV color features to HOG features makes a remarkable improvement in comparison with HOG features only. At a FPPW of 10^{-4}, a usually used reference point, the classifier with LUV features shows a miss rate of 0.4 that is much lower than 0.54 of the classifier with no LUV features. Color information cannot be competent if it is used alone, but the result reveals that color can be a complementary, effective source when combined with HOG. Conflicting results have been reported on whether such raw color features are beneficial [11],[12]. Our results confirm the fact that raw color is salient information that enables us to reject head-shoulder-free images much easier than no color information. In fact, detection rate of 0.6 at the FPPW of 10^{-4} is worse than the detection rate of 0.9 that SVM-based full-body detectors of [2][11] reported for the INRIA dataset, which implies the fact that head-shoulder detection is harder than full-body detection.

When the GSS features were added into the HOG and color feature set, it made the performance improvement by reduction of 12% missing rate at a FPPW of 10^{-4} than with the HOG+LUV features. Being considered a marginal improvement in comparison with the improvement in case of utilization of color information, decrease in the missing rate larger than 10% at a FPPW of 10^{-4} is a significant improvement. The HOG GSS features were chosen more frequently in training than the color GSS features.

We performed another experiment to examine the effectiveness of the GSS features depending on human pose, i.e., angle view. The test dataset was split into three separate datasets of the head-shoulders for frontal view, rear view, and profile view. Classifiers with the GSS features were tested with the three datasets, and the results are shown in Fig. 8. As expected, it is remarkably observed that performance in the frontal-view dataset, the missing rate of 0.3 at the FPPW of 10^{-4} is significantly better than that in the profile-view dataset, the missing rate of about 0.45. It should be noted that Fig. 8(a)-(b)reveals that the GSS features contribute to better performance for the frontal-view dataset than in the profile-view one. It is reasonable that the GSS features are effective for the frontal-view head-shoulder detection as they have been developed based on the self-similarity or vertical symmetry of head-shoulder patterns.

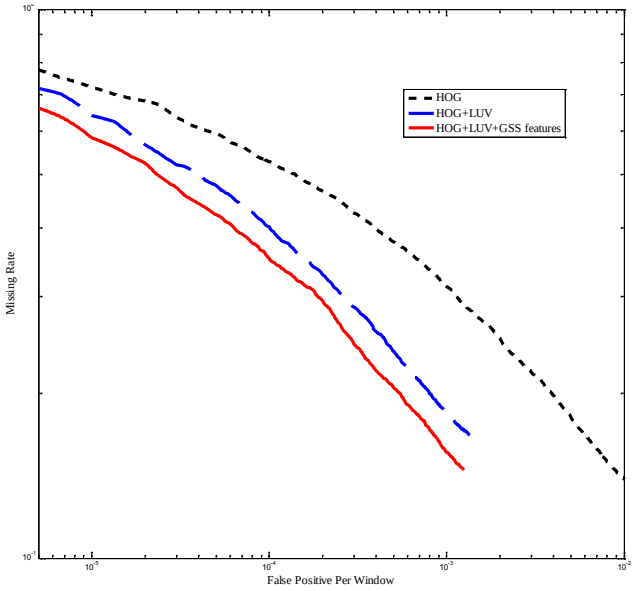
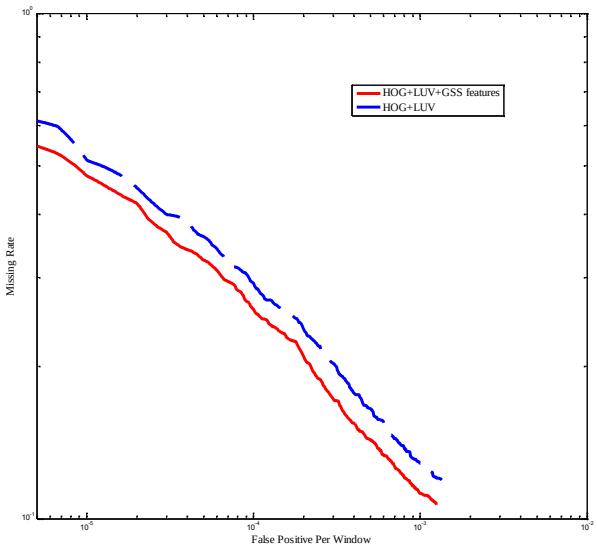
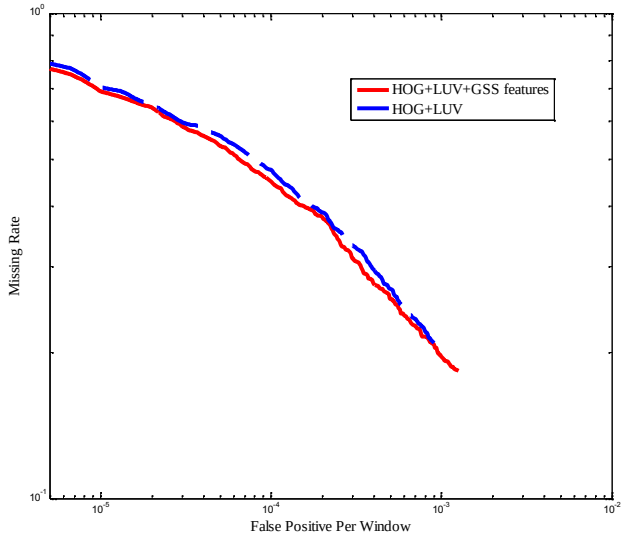


Fig. 7 Performance of classifiers with HOG, HOG+LUV, and HOG+LUV+GSS features



(a)



(b)

Fig. 8 Evaluation of the GSS features. (a) Performance on the frontal-view dataset. (b) Performance on the profile-view dataset

Some detection results of our head-shoulder detector on videos captured by a goggle-typed camera while walking a downtown street are shown in Fig. 9. Solid green rectangles and dotted white ones represent positives and false positives, respectively. It can be seen that a majority of false positives occur around torso and either of shoulders. The detection speed of our detector is about 5-10 frames per second at 20 detection scales for a 320x240-sized image on an Intel 2.7GHz CPU.

Fig. 9 Detection results on a downtown street

V. Conclusions

In this paper, we have introduced new features that exploit global self-similarity (GSS) of color and gradient information in head-shoulder detection in crowded scenes. In a cascade-correlation classifier, a variant of Viola-Jones's cascade classification framework, the proposed features have been evaluated on our own head-shoulder dataset. Experimental results have shown that the GSS features were effective in reduction of false alarms when used in combination with local features, e.g., HOG and color information.

Acknowledgment

This work was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (No. 2012K001342).

References

[1] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection:Survey and experiments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2179–2195,2009.
[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey on pedestrian detection for advanced driver assistance systems," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1239–1258, 2010.
[3] K-Y K. Wong, R. H. Y. Chung, Y. L. Francis, and K-P P. Chow, "Real-time multiple head shape detection and tracking system with decentralized trackers," in *Proc. Intelligent Systems Design and Application*, 2006, pp. 384–389.
[4] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet part detectors," *International Journal of Computer Visions*, vol. 75, pp. 247–266, 2007.
[5] M. Li, Z. Zhang, K. Huang, and T. Tan, "Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection," in*Proc. Int. Conf. on Pattern Recognition*, 2008. pp. 1–4.
[6] X. Ding, H. Xu, P. Cui, L. Sun, and S. Yang, "A cascade SVM approach for head-shoulder detection using histograms of oriented gradients,"in *Proc. IEEE Int. Symposium on Circuits and Systems*, 2009, pp. 1791–1794.
[7] C. Zeng and H. Ma, "Robust head-shoulder detection by PCA-based multilevel HOG-LBP detector for people counting,"in *Proc. Conf. on Pattern Recognition*, 2010, pp. 2069–2072.
[8] N. Dalal and B. Triggs,"Histograms of oriented gradients for humandetection,"in*Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
[9] P. Viola and M. J. Jones, "Rapid objection detection using a boosted cascade of simple features," in*Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.
[10] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in*Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2006, pp. 1491–1498.
[11] P. Dollar, Z.Tu, P. Perona, and S. Belongie, "Integral channel features,"in*Proc. British Machine Vision Conference*, 2009.
[12] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2010, pp. 1030–1037.
[13] C-H Chuang, S-S Huang, L-C Fu, and P-Y Hsiao, "Monocular multi-human detection using augmented histograms of oriented gradients,"in*Proc. Int. Conf. on Pattern Recognition*, 2008. pp. 1–4.
[14] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," in *Advances in Neural Information Processing Systems 2*, Margna-Kaufmann, 1990, pp. 524–532.
[15] C. Huang, H. Ai, B. Wu, and S. Lao, "Boosting nested cascade detector for multi-view face detection," in *Proc. Int. Conf. on Pattern Recognition*, 2004. pp. 415–418.