Reduction of False Positives in Head-Shoulder Detection Based on Multi-Part Color Segmentation

Lae-Jeong Park

Abstract—The paper presents a method that utilizes figure-ground color segmentation to extract effective global feature in terms of false positive reduction in the head-shoulder detection. Conventional detectors that rely on local features such as HOG due to real-time operation suffer from false positives. Color cue in an input image provides salient information on a global characteristic which is necessary to alleviate the false positives of the local feature based detectors. An effective approach that uses figure-ground color segmentation has been presented in an effort to reduce the false positives in object detection. In this paper, an extended version of the approach is presented that adopts separate multipart foregrounds instead of a single prior foreground and performs the figure-ground color segmentation with each of the foregrounds. The multipart foregrounds include the parts of the head-shoulder shape and additional auxiliary foregrounds being optimized by a search algorithm. A classifier is constructed with the feature that consists of a set of the multiple resulting segmentations. Experimental results show that the presented method can discriminate more false positive than the single prior shape-based classifier as well as detectors with the local features. The improvement is possible because the presented approach can reduce the false positives that have the same colors in the head and shoulder foregrounds.

Keywords—Pedestrian detection, color segmentation, false positives, feature extraction.

I. INTRODUCTION

PEDESTRIAN detection from images is one of the challenging tasks in computer vision because of complex backgrounds and variations of human poses [1]. Detecting pedestrian in crowded environments such as shopping malls is more difficult due to frequent occlusion. Some researches on upper body detection such as head and shoulders have been made to detect humans in those circumstances [2]-[7]. In particular, head-shoulder pattern detection has attracted attentions as the pattern has more salient, informative shape than the head part only. Most of the pedestrian detectors adopt local and rapid-to-compute features such as HOG [8] and SIFT [9] because of the real-time requirement. However, the detectors sometimes produce false positives (an image is classified as one having a head-shoulder pattern even though it does not) for input images that are not be similar to the head-shoulder pattern in terms of human perception, as shown in Fig. 2.

L.-J. Park is 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).

There has been some research on the features that are capable of supplementing the limitation of the local features so that the false positives could be alleviated [10], [11]. They were built based on other "global" shape cues than the "local" edge-like features. Among the cues, effectiveness of "color" cue has been evaluated compared with the local features in terms of reduction of the false positives. Also new approaches and features have been proposed that utilizes color cues explicitly or implicitly to extract global characteristics in an image [11]-[15]. Diwala et. al. performed an empirical study of contextual cues including color information in object detection [12]. Ott and Everingham proposed a new feature that encodes color information in each local block implicitly and is computable readily during HOG extraction [13]. On the other hand, Walk et. al. introduced a so-called color self-similarity (CSS) feature that represents global distribution of colors by using covariances between color histograms of local blocks and evaluated its performance on pedestrian detection [11]. Ramanan suggested a new approach that can verify false positives by utilizing figure-ground color segmentation based prior shapes [14]. Park introduced an extension of Ramanan's method by leveraging the figure-ground color segmentation in a manner that separate color segmentation is performed with multi prior shapes and evaluated the performance in the pedestrian upper body detection [15].

This paper generalizes the extended version of the color segmentation-based approach by incorporating more new general shaped foregrounds than in [15]. Color segmentation with the additional multi-part foregrounds enables us to extract the features that could be more effective in reducing the false positives than in just the prior shapes in [14], [15]. In addition to reduction of some of the false positives where multi parts share the same colors, the proposed approach make it possible to eliminate some of the false positives in which there are the same colors around the outline of the head-shoulder shape.

The remainder of the paper is organized as follows. Section II describes the extraction of the global feature based on color segmentation with prior shapes and multipart shapes in detail. Section III evaluates the performance of the proposed approach with the experimental results on a head-shoulder dataset and conclusions are made in Section IV.

II. EXPLOITING GLOBAL FEATURE BASED ON FIGURE-GROUND COLOR SEGMENTATION

In this section, the extraction of global features based on figure-ground color segmentation is described in detail. With a foreground shape on an input image, figure-ground segmentation is computed on the basis of the color histograms

of the inside and the outside of the foreground region. The segmentation results are used as the global feature. The feature can be regarded as being "global" in the sense that color information of both the inside and the outside of the foreground region are collected in the segmentation computation.

A. Single Shape-Based Color Segmentation

A prior shape-based color segmentation [14] is described briefly for readers' convenience. Fig. 1 illustrates the process of the figure-ground color segmentation and its segmentation result.
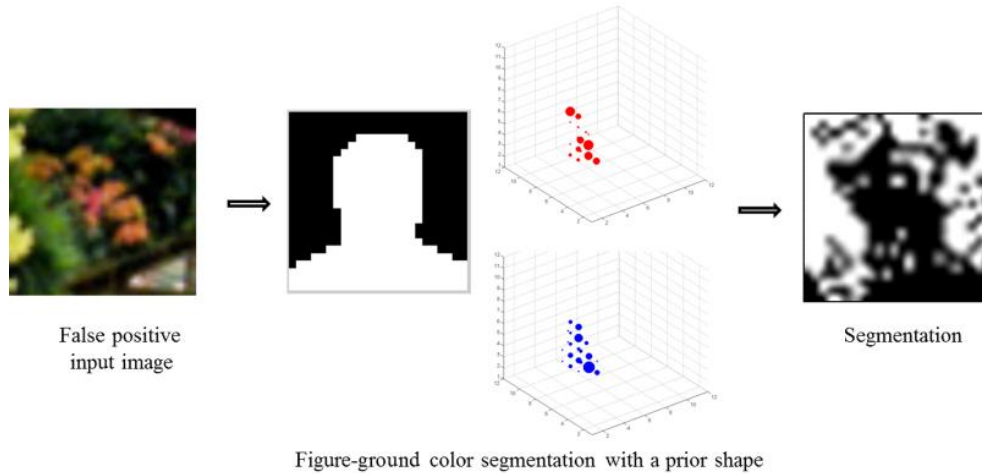

Fig. 1 The color segmentation with a prior single shape in the head-shoulder detection.

To be specific, the foreground region and the background region are first divided by a prior shape which is dependent on the object we want to detect. A head-shoulder shape is used in head-shoulder detection as shown in Fig. 1. Then, two color histograms are computed for the foreground and the background respectively. Color images can be represented by 3D color spaces such as RGB and HSV, but HSV color space is preferred with consideration of previous studies in [7]. Given the two color histograms, figure-ground segmentation is computed by a graph-cut [16] that is performed by optimization of a potential function which is given by

$$E(l_1, l_2, ..., l_K) = \sum_k c(l_k) + \mathsf{r} \sum_{j,k \in N} I(l_j \neq l_k),\qquad(1)$$

where l_k represents a label (either foreground or background) for a pixel at the location k, $c(l_k)$ represents a label cost function that determines to what extent a pixel favors the foreground label versus the background label, N is the set of 4-connected neighbors, $I()$ represents a penalty function for the mismatch of labels between pixels in the neighborhood, and r is a weighting coefficient. The cost function $c(l_k)$ is given by

$$c(l_k = FG) \propto -\log[H_{FG}(im(k))],\qquad(2)$$
$$c(l_k = BG) \propto -\log[H_{BG}(im(k))],$$

where H_{FG} (H_{BG}) represents a color histogram of the foreground (background) and $im(k)$ represents a bin number of the color histogram for pixel at the location k. Therefore, $H_{FG}(im(k))$ indicates the number of pixels that fall into the bin of the color histogram of the foreground region. In optimization based on (1), (2), a pixel at the location k gets higher chance to be assigned to the foreground if there are more pixels that have the same color as the pixel at the location k in the color histogram of the foreground than in the color histogram of the background. It should be noted that two colors are regarded as the "same" color only when they fall into the same bin in the binned color space.

B. Extended Multipart Shapes-Based Color Segmentation

The feature extracted through the color segmentation with a prior shape has contributed to reduction of false positives which have the same color in both the inside and the outside of the shape, i.e., the head-shoulder shape and the background [14]. An input image, either of the true positive or the false positive, is more likely to be a false positive if the same colors are distributed across the foreground and the background. For the input images with such color distribution, the color segmentation process of Fig. 1 generates different segmentations from the shape of the foreground so that the segmentation results can be used to distinguish the true positive from the false positive. Specifically, for the input image in which the foreground and the background share the same colors, H_{FG} and H_{BG} have nonzero values at the bins corresponding to the colors. The pixels having the same color will be assigned to either the foreground or the background in the minimization of the potential function of (1) in such a way that they are more likely to belong to the foreground if H_{FG} at the bin is larger than H_{BG}. That is, for the color that the foreground and the background share, the ratio of the numbers of pixels in the two color histograms plays a major role in determining the label.

Fig. 2 Examples of undesirable segmentations of some false positives in the head-shoulder detection by a single shape-based color segmentation

Unfortunately, it happens not infrequently that even though an input image is a false positive (there is no foreground shape in the image), the figure-ground segmentation leads to a segmentation that is seemingly similar to the shape of the desired foreground. Fig. 2 shows the undesirable segmentations of some false positives in the head-shoulder detection. There can be two cases. The first is the case that segmentation similar to the shape of the foreground is produced because the foreground and the background never have the same colors. The second is the case that, even though parts of the foreground share the same colors as parts of the background, the color segmentation process results in an unwanted segmentation that is closely similar to the shape of the foreground because either of the numbers of the pixels having the same colors in the both regions are quite small.
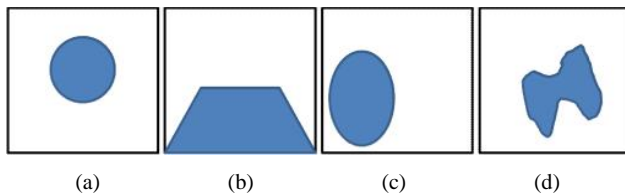

Fig. 3 Candidates of the foregrounds in the head-shoulder detection. (a) a head foreground, (b) a shoulder foreground, (c) a foreground near the left shoulder, (d) an arbitrary foreground

In order to reduce such false positives, we propose a new color segmentation approach that adopts separate multiple foregrounds instead of a single prior foreground of [14] and that performs the color segmentation process repeatedly by using each of the foregrounds. An object usually has several distinct parts, each of which tends to have different colors from the others. For example, in the pedestrian upper body detection, the object has different colors in its parts such as hair, skin, and clothing. Figs. 3 (a), (b) show an example of two candidates of the multiple foregrounds in the head-shoulder detection, respectively (a head part and a shoulder part). In general, it is desirable that the multiple foregrounds are chosen by taking consideration of common characteristics of colors in the distinct parts of the object to detect. On the contrary to this, some of the false positives have the same colors across the

distinct parts of the object, e.g., the head part and shoulder part. Hence, by adopting the distinct foregrounds that correspond to the parts and usually have different colors from each other we can maximize the opportunity of classifying the false positives having the same colors in the both distinct foregrounds. To be specific with reference to Figs. 4 (a), (b), the color segmentation process with the head part foreground and the shoulder part foreground produces two distinct segmentations. When the same colors distribute across the head part and the shoulder part in an input false positive image, either of the segmentation results does not match up with its own foreground shape, making it possible to classify the false positive with the segmentation result. For example, if the head part and the shoulder part have the same colors, no head shape is highly likely to appear in the segmentation with the head part foreground. The reason for this is that, because the head part is smaller in area than the shoulder part, ratios of H_{FG} to H_{BG} at the bins corresponding to the colors is much smaller in the head-shaped foreground than in the single head-shoulder foreground, leading to high possibility of no appearance of the head-shaped segmentation. Here, it should be noted that the ratio of H_{FG} to H_{BG} is dependent upon the shape and area of the foreground chosen. The two distinct segmentation results are concatenated and used as features to a classifier.

Beside the introduction of the distinct foregrounds constituting parts of the object, additional arbitrarily shaped foregrounds are added into the color segmentation process for further improvement. Through the investigation on the false positives that occur even with the two distinct foregrounds, we found that some false positives have the same colors around the boundary between the head-shoulder foreground and the background. Generally, the head part and the shoulder part in pedestrian's upper-body do not have the same colors as the background. Therefore, when the color segmentation process is performed with a new foreground that lies somewhere around the outline between the target object and the background, e.g., a foreground in Fig. 3 (c), the resulting segmentation is more likely to be similar to the shape of the new foreground if colors inside the foreground are homogeneous than the inhomogeneous color distribution. The resulting segmentation depends not only on the size of position of the additional foreground but also on colors in the outside of the new foreground. However, for a foreground locating around the outline of the target object, it is, on average, reasonable to determine if an input image is false positive based on to what degree the shape of the foreground appears in the resulting segmentation. The foregrounds added can be either carefully crafted manually or chosen by optimization methods. Fig. 4 illustrates a complete process of the proposed color segmentation scheme.

Given each of the figure-ground color segmentation result (0 or 1), it is smoothed by a 3x3 sliding window and is then vectorized. The resulting feature is constructed by concatenating the vectors and is used as inputs to a classifier.
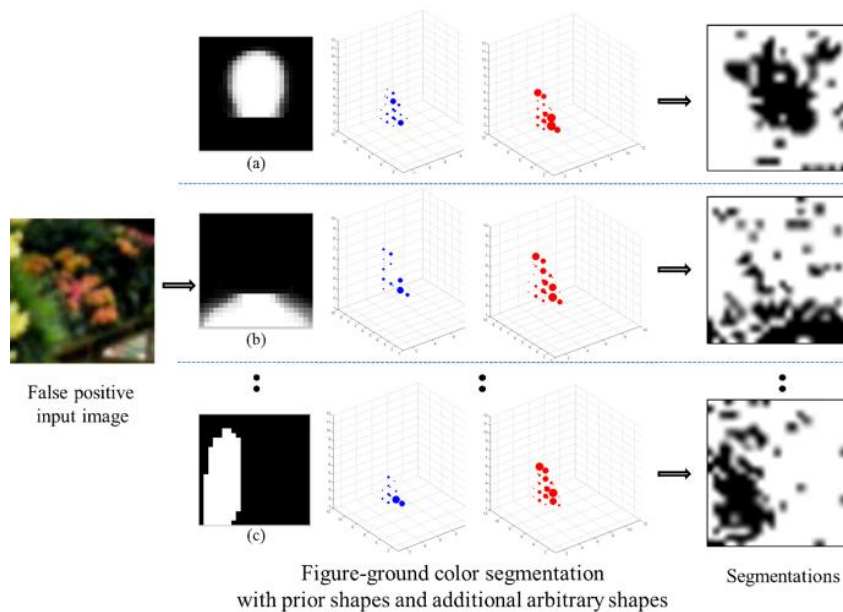
Fig. 4 The presented approach based on color segmentation with different multiple foregrounds in the head-shoulder detection

III. EXPERIMENTAL RESULTS

A. Dataset

To evaluate the presented approach in the head-shoulder detection, a dataset was constructed by cropping the near-frontal pedestrian images from a public INRIA dataset [8] and from the Internet and adding many negative images collected from the Internet into the INRIA negative image set. Negative images of shopping mall, subway platform, and the downtown were included since the places are environments where the head-shoulder detector operates. The dataset consists of 757 positive samples (near-frontal head-shoulder images) and 1,250 negative samples. The negative samples consist of the false positive images that passed through a head-shoulder detector [7] that uses HOG-like local features among a large number of patches of the negative image set. The dataset of 24x24-sized images is divided into a training dataset (75%) and a test dataset (25%).

B. Experimental Conditions

The figure-ground color segmentation was performed in not RGB color space but HSV color space. We have found from the preliminary results in [15] that HSV color space provided better performance than the RGB and the CIE color spaces. Non-uniform bin sizes were adopted such that bin sizes for the H axis and the other axes were set to 12 and 24, respectively.

Simulated annealing (SA) [17], which is known as an effective search algorithm for the global optimization, was chosen to search the best candidates of the foregrounds added. For simplicity and computation time, rectangular shapes only are considered as the foregrounds in this experiment. The search space of SA is the coordinates of the lower-left and upper-right points of a rectangle. The perturbation is made on the coordinates of a rectangle by adding Gaussian nose $N(0,\dagger)$ where $\dagger$ is set to 3. The value to be maximized is an

area under ROC (receiver operating characteristic) curve, so-called AUC (area under curve). The AUC is a common evaluation metric for two-class classification and the AUC for a perfect classifier is equal to 1. The initial and final values of the temperature, which are a major parameter of SA, are set to 0.05 and 0.002, respectively.

When evaluating the features extracted based on the color segmentation process, we use SVM (support vector machine) with the linear kernel rather than that with the nonlinear kernel. The linear kernel is generally worse than the nonlinear one, but it offers a good environment where we compare the effectiveness of the features fairly. For each feature, twenty runs were performed with different random sets of the training and test datasets and their results were then averaged.

C. Results

Performance is evaluated in terms of ROC curve and AUC. Fig. 5 shows the ROC curves of SVMs with features extracted from the single head-shoulder foreground (dotted line), from the two head and should foregrounds (dash-dotted line), from the two foregrounds plus a single foreground added (dashed line), and from the two foregrounds plus two foregrounds added (solid line). Their AUC values are shown in Table I. The feature extracted from color segmentation with the single head-shoulder foreground outperforms the HOG feature (its AUC is 0.85 and the ROC curve is not shown in Fig. 5). The superiority of the feature extracted from the color segmentation approach implies that it is hard for local features such as HOG to represent global characteristic such as color distribution on an input image. As can be seen in Fig. 5 and Table I, the partitioning the single head-shoulder foreground into the two distinct foregrounds, or head part and shoulder part foregrounds leads to better performance by 0.016 in terms of AUC than the single prior of [14]. Incorporation of more foregrounds into the

two head and shoulder foregrounds makes remarkable improvements. Adding a single rectangular foreground found by SA makes an improvement by 0.029 in terms of AUC over the two head and shoulder foregrounds. One more foreground results in further improvement by 0.004 than the three multipart foregrounds. The sets of the best foregrounds found by SA are shown in Fig. 6. It is noteworthy that the foregrounds added include parts of the outline of the head-shoulder pattern, i.e., the part of neck region and the right part of the pattern to detect, as expected.
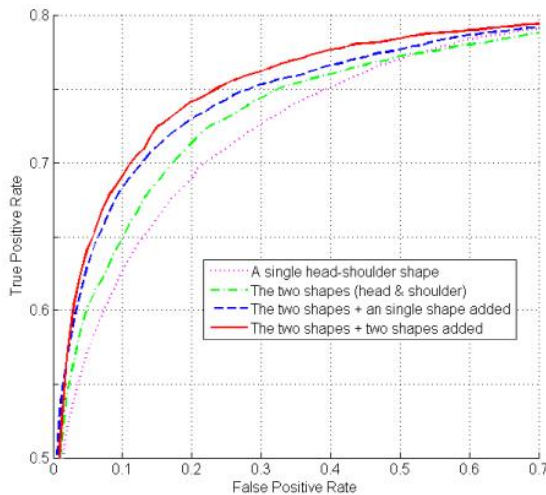

Fig. 5 ROC curves of SVM classifiers with different sets of the features based on color segmentation

TABLE I
PERFORMANCE OF SVMS WITH THE FEATURES IN TERMS OF AUC

Head-shoulder foreground	Head & shoulder foregrounds	Head, shoulder, & single optimized foregrounds	Head, shoulder, & two optimized foregrounds
0.874	0.890	0.919	0.923

Another experiment was performed to examine if the more foregrounds are added, the better performance is. Fig. 7 shows change of AUCs with respect to the number of foregrounds added. Interestingly, one and two foregrounds contribute to the improvement, but the performance gets worse slightly as the number of the foregrounds added increases. It seems to be caused by the overfitting due to the increase of the dimension of the feature as the number of the foregrounds added increases.
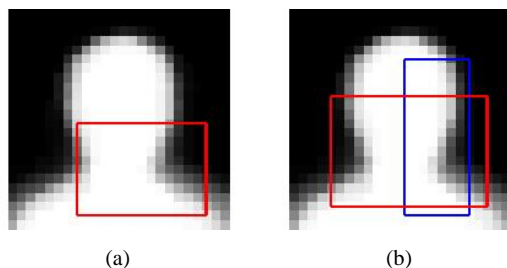

Fig. 6 The sets of foregrounds found by SA optimization when (a) a single foreground is added and (b) two foregrounds are added.
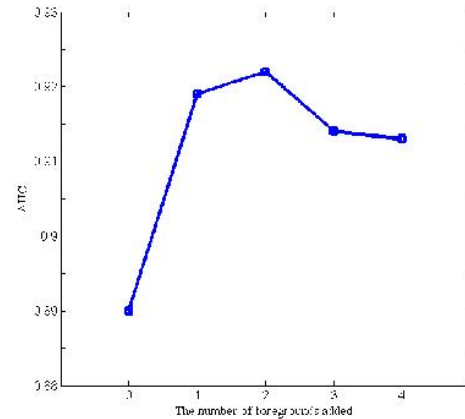

Fig. 7 Change of AUCs with respect to the number of foregrounds added; no foreground represents the two foregrounds of head and shoulder only

IV. CONCLUSIONS

This paper has suggested a new approach that maximizes the usefulness of figure-ground color segmentation in order to extract global feature capable of reducing the false positives in head-shoulder detection that would be missed with local features. The proposed approach extends single prior shape-based figure-ground color segmentation by introducing separate multipart foregrounds and separately performing the figure-ground color segmentation with each of the separate foregrounds. The set of the multipart foregrounds consists of not only the parts of the head-shoulder shaped foreground, i.e., two shapes of head and shoulder but also auxiliary foregrounds that are optimized by a search algorithm. A set of the resulting segmentations is used as the feature that enables us to reduce the false positives that cannot be eliminated with a single prior-based approach.

Experimental results on a head-shoulder dataset have shown that the presented approach was more effective in the reduction of the false positive than that with the single prior foreground. The presented approach made an improvement by 5.5~6.5% in terms of the true positive at a range of the false positive of 10~20%. The improvement is attributed to the reduction of the false positives that have the same colors in the head and shoulder foregrounds. As expected in the introduction of the auxiliary foregrounds, the foregrounds found by a search algorithm contain parts of the outline of the head-shoulder pattern in which colors could be uniform for some of the false positives. Adaptive binning for the color histogram and arbitrarily shaped foregrounds are under study for further improvement.

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] 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.

[3] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet part detectors," *Int. J. of Computer Visions*, vol. 75, pp. 247–266, 2007.

[4] 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.

[5] 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.

[6] 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.

[7] L.-J. Park and J-H Moon, "Exploiting global self-similarity for head-shoulder detection," *World Academy of Science, Engineering, and Technology*, vol. 7, no. 4, pp. 144–148, 2013.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[9] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Conf. on Computer Vision*, 1999, pp. 1150–1157.

[10] P. Dollar, Z.Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. British Machine Vision Conference*, 2009,

[11] 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.

[12] S. K. Diwala, D. Hoiem, J. H. Hays, A. A. Efros, and M. Hebert, "An empirical study of context in object detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2009, pp. 1271–1278.

[13] P. Ott and M Everingham, "Implicit color segmentation features for pedestrian and object detection," in *Proc. IEEE Conf. on Computer Vision*, 2009, pp. 723–730.

[14] D. Ramanan, "Using segmentation to verify object hypothesis," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[15] L.-J. Park, "Exploiting color segmentation in pedestrian upper-body detection," *J. of the institute of Electronics and Information Engineers*, vol. 51, no. 11, pp. 181–186, 2014 (in Korean).

[16] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," in *Proc. IEEE Conf. on Computer Vision*, 1999, pp. 374–384.

[17] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.