Pattern Recognition Letters 42 (2014) 148–153

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

An improved Haar-like feature for efficient object detection ☆

Ki-Yeong Park, Sun-Young Hwang *

Department of Electronic Engineering, Sogang University, C.P.O. Box 1142, Seoul 100-611, Republic of Korea

ARTICLE INFO

Article history:
Received 24 August 2013
Available online 3 March 2014

Keywords:
Haar-like feature
Illumination-invariant feature descriptor
Classifier
Object detection
Real-time system

ABSTRACT

In this paper, we propose an improved feature descriptor, *Haar Contrast Feature*, for efficient object detection under various illumination conditions. The proposed feature uses the same prototypes of Haar-like feature and computes contrast using the normalization factor devised to reflect the average intensity of feature region. It is computed efficiently using an integral image and is more powerful in real-time applications by not requiring variance normalization during detection process. It shows improved performance under a wide range of illumination conditions. For experiments, classifiers for face, pedestrian, and vehicle were trained by employing the conventional Haar-like feature with/without variance normalization, the local binary pattern descriptor, and the proposed feature descriptor, and their performances were evaluated. Experimental results confirm that classifiers employing the proposed feature descriptor outperform those employing the conventional Haar-like feature or the local binary pattern descriptor in terms of detection accuracy under most illumination conditions.

1. Introduction

Object detection is a classification problem which distinguishes a class of objects such as faces, pedestrians, and vehicles from cluttered backgrounds, and has applications in various areas including image retrieval, video surveillance, and advanced driver assistance systems. One of the major issues in object detection is how to represent objects. Contrary to the traditional pattern classification problem where decision is made between well-defined classes, objects are various in color and texture and backgrounds against which objects lie are unconstrained. Object representation must accommodate the intra-class variability without compromising discriminative power in distinguishing objects from cluttered backgrounds [1].

Haar-like feature has been popularly employed for object representation. Papageorgiou et al. [1] proposed to represent objects in terms of a subset of over-complete dictionary of Haar wavelet basis functions, and Viola and Jones [2] developed the Haar-like feature by extending the idea of using Haar wavelets. Viola and Jones introduced fast object detector based on the Haar-like feature. A new image representation called integral image was introduced to compute Haar-like features efficiently, and AdaBoost [3,4] was used both to select a small number of most discriminating features and to train classifiers. Classifiers were combined in a cascade which allows background regions to be quickly discarded while spending more computation on object-like regions [2]. Object detector based on Haar-like feature has been very successful due to its efficiency in the domain of face detection [5,6,7], and has also been adopted for pedestrian detection [8,9,10,11,12] and vehicle detection [13,14,15]. Enzweiler and Gavrila [16] performed experiments to compare several pedestrian detectors and reported that the pedestrian detector based on Haar-like feature is superior at lower image resolution and shows near real-time processing speeds, while the pedestrian detector based on histogram of oriented gradients (HOG) has a clear advantage at higher image resolutions but with lower processing speeds. They also reported that the average processing time of the pedestrian detector based on Haar-like feature was approximately 20 times faster than that of other detectors. Dollár et al. [17] evaluated performance of 16 state-of-the-art pedestrian detectors and reported that a combination of Haar-like features, shapelets, shape context, and HOG features [18] outperforms other detectors with channel feature [19] a close second. Channel feature computes Haar-like feature over multiple channels of visual data. In spite of this success, Haar-like feature has a fundamental limitation in that it is not illumination-invariant.

Detection performance is significantly degraded under variable lighting. Problems related to illumination variations have been investigated by several researchers, mostly in the domain of face detection and recognition. One of those approaches is to use illumination models [20,21,22]. Illumination models are used for learning the variations caused by illumination changes. This approach is not appropriate for real-time applications, since it requires a lot of

☆ This paper has been recommended for acceptance by Dr. J. Yang.

* Corresponding author. Tel./fax: +82 2 703 0582.
E-mail address: hwang@sogang.ac.kr (S.-Y. Hwang).

http://dx.doi.org/10.1016/j.patrec.2014.02.015
0167-8655/

computations for reconstruction, rendering, and training. Cordiner et al. [23] proposed the use of multiple classifiers, where several face detectors are separately trained for different illumination environments. The other approach is to normalize images for reducing the effect of illumination variations, where normalization is performed in preprocessing stage to re-illuminate images into a canonical illumination environment [24]. Another approach is to use illumination-invariant features which maintain stable values under different illumination conditions. Illumination-invariant features are usually derived from high-frequency spatial images [25]. These approaches achieved illumination invariance to a certain degree, but with a lot of computations.

In this paper, we propose an improved feature descriptor for efficient real-time object detection under a wide range of illumination conditions. The proposed feature computes contrast using the normalization factor devised to reflect the average intensity of feature region. The rest of this paper is organized as follows: In Section 2, we present a brief introduction of the conventional Haar-like feature. The proposed feature is presented in Section 3. Experimental results are presented in Section 4, and the conclusions are drawn in Section 5.

2. Backgrounds

In this section, a brief introduction of the conventional Haar-like feature is presented. Haar-like feature is not illumination-invariant and image sub-windows have to be variance normalized to minimize the effect of illumination variation during training and detection processes.

2.1. Haar-like feature

Haar-like feature was motivated by the observation that while absolute intensity values of different regions change dramatically under varying illumination conditions, their mutual ordinal relationship remains largely unaffected [26]. A Haar-like feature consists of two or more vertically or horizontally adjacent rectangular regions, and its value is difference between sums of pixels within these rectangular regions [2]. Value of a Haar-like feature (HF) can be represented as weighted sum of intensities of rectangular regions in a feature as in (1).

$$\mathrm{HF} = \sum_{i=1}^{R} sign(i) \cdot w_i \cdot \mu_i \tag{1}$$

where R is the number of rectangular regions constituting a Haar-like feature and $sign(i)$ is the sign assigned to the ith rectangular region. w_i is the weight which is inversely proportional to the area of the rectangular region and μ_i is the average intensity of the rectangular region. Average intensity of a rectangular region can be computed efficiently in constant time at any scale using an integral image [2]. A pixel of an integral image at (x,y) has sum of pixels within the rectangular region from $(0,0)$ to (x,y) in the original image. It requires at least 2 mn additions to generate an integral image for an $m \times n$ image [27]. After generating an integral image, sum of pixels within any rectangular region is computed by referencing only 4 pixels from the integral image.

2.2. Probability distribution of Haar-like feature value

AdaBoost [3,4] is used both to select features and to train a classifier [2]. At each AdaBoost round, threshold is determined for each feature, and a feature showing the lowest expected error for given training samples is selected from over-complete set of features and used as a weak learner [28].

Haar-like features selected by AdaBoost show distinct probability distributions. Fig. 1 shows probability distributions of typical Haar-like features. Since it is highly possible for two adjacent regions to have similar intensity in a randomly selected image sub-window, probability distribution for negative samples forms a peak around zero. On the other hand, probability distribution for positive samples can form a peak at any value. Threshold for a feature will be set to a value between these two peaks, and the expected error of the feature is calculated with the threshold. The feature in Fig. 1(a) has low expected error, and the feature in Fig. 1(b) has high expected error. Since probability distribution for negatives forms a peak around zero, the probability that a feature is determined to 'negative' increases as value of the feature becomes smaller.

2.3. Variance normalization for Haar-like feature

To minimize the effect of illumination variation, Viola and Jones [2] variance normalized image sub-windows during training and detection processes. Variance is computed as $\sigma_w^2 = \mu_w^2 - 1/a_w \cdot \Sigma x^2$, where σ_w and μ_w are the standard deviation and the average of pixel values within an image sub-window, respectively. a_w is the area of the image sub-window and x is value of a pixel within the image sub-window. While μ_w can be computed using the integral image prepared for computation of Haar-like features, an additional integral image of squared pixels is required to compute Σx^2. Once variance of pixels within an image sub-window is computed, variance-normalized value of a Haar-like feature (NHF) is computed by post-multiplying feature values rather than pre-multiplying pixels within the image sub-window [2] as in (2).

$$\mathrm{NHF} = \frac{1}{\sigma_w} \cdot \sum_{i=1}^{R} sign(i) \cdot w_i \cdot \mu_i \tag{2}$$

3. The proposed Haar Contrast Feature

In this section, we propose an improved feature descriptor for efficient real-time object detection under a wide range of illumination conditions.

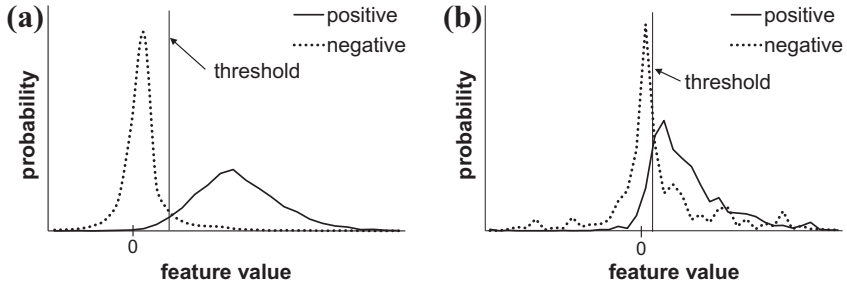

Fig. 1. Probability distributions of typical Haar-like features. (a) Feature with low expected error, (b) feature with high expected error.

3.1. Consideration on intensity difference under illumination variations

Contrast and brightness of images vary under different illuminations. Variations in contrast and brightness result in changes in intensity difference between adjacent regions. The intensity difference between adjacent regions corresponds to the value of a Haar-like feature. Contrast variation and brightness variation in a specific region can be represented by variance of pixels within the region and average intensity of the region, respectively. To minimize the effect of contrast variation on the value of a Haar-like feature, variance normalization can be used [2]. However, the effect of variance normalization is limited due to weak correspondence between the variance of pixels within an image sub-window and the feature value computed from a small part of the image sub-window.

As value of a Haar-like feature can be normalized by variance of pixels within an image sub-window, it can be normalized by exploiting its correspondence to average intensity of the image sub-window or that of the feature region. Fig. 2 shows values of selected Haar-like features computed for both positive and negative samples. Values of first 5 features of a vehicle classifier are computed for each training sample. Feature values vs. the standard deviation of pixel values within a sample, those vs. the average intensity of a sample, and those vs. the average intensity of a feature region are shown in Fig. 2(a)–(c), respectively. Feature values show stronger correspondence to the average intensity of a feature region than to the standard deviation of pixel values within a sample or to the average intensity of a sample. While feature values are distributed in a small range for some features and spread for other features when the standard deviation of pixel values within a sample is very large in Fig. 2(a) or when the average intensity of a sample is very small or very large in Fig. 2(b), they are always distributed in a small range when the average intensity of feature region is very small or very large in Fig. 2(c). Utilizing the above observation, we can make feature values fluctuate less under various illumination conditions by normalizing them with average intensity of feature region.

3.2. Definition of Haar Contrast Feature

The proposed Haar Contrast Feature computes contrast of feature region instead of intensity difference between rectangular regions within a feature. Contrast between adjacent regions can be defined as intensity difference between the regions divided by their average intensity. Intensity difference between rectangular regions within a feature may decrease as illumination becomes very low or very high as shown in Fig. 2(c). To make a feature illumination-invariant, its value needs to be increased when the average intensity of the feature region is very low or very high. For this, normalization factor is devised to have small values at those regions. Fig. 3 illustrates normalization of feature value with

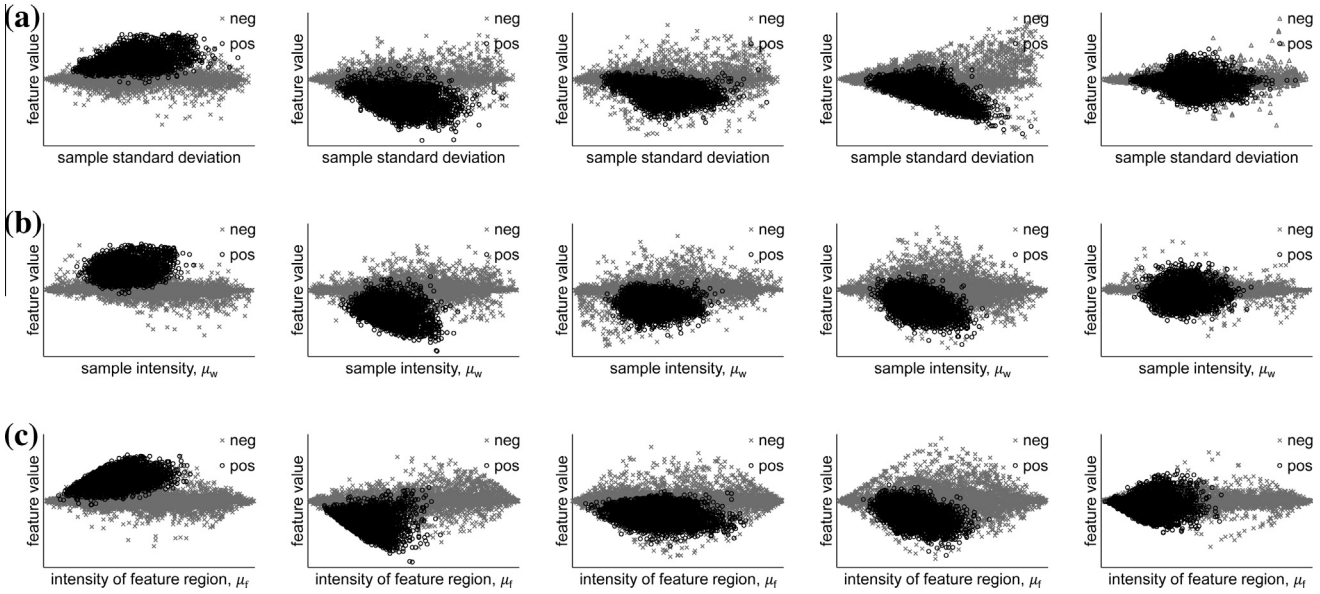

Fig. 2. Values of first 5 features out of Haar-like features used for a vehicle classifier. (a) Feature value vs. standard deviation of pixel values within a sample, (b) feature value vs. average intensity of a sample, (c) feature value vs. average intensity of feature region.

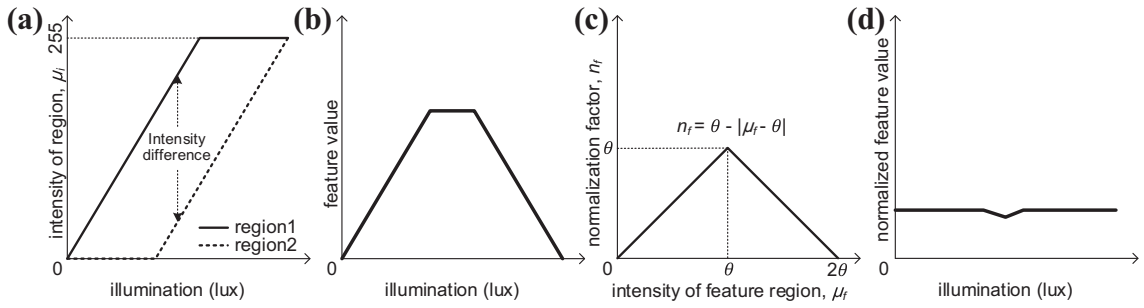

Fig. 3. Normalization of feature value. (a) Intensity difference between two rectangular regions under different illuminations, (b) feature value before normalization, (c) normalization factor, (d) normalized feature value.

the devised normalization factor. Fig. 3(a) shows intensity difference between two rectangular regions under different illuminations, and Fig. 3(b) shows feature value before normalization. Feature value decreases as the average intensity of those regions becomes very low or very high. Fig. 3(c) shows the normalization factor whose value increases when the average intensity is less than θ and decreases afterwards. Normalized feature value is shown in Fig. 3(d). The normalization factor n_f is defined as in (3).

$$n_f = \theta - |\mu_f - \theta| \tag{3}$$

where μ_f is the average intensity of a feature region and θ is threshold. The proposed feature is defined in the range of $0 < \mu_f < 2\theta$, and it returns zero without computation when $\mu_f = 0$ or $\mu_f \geqslant 2\theta$ as in (4).

$$\mathrm{HCF} = \begin{cases} \frac{1}{n_f} \cdot \sum_{i=1}^{R} sign(i) \cdot w_i \cdot \mu_i, & \text{if} \quad 0 < \mu_f < 2\theta \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

It seems reasonable that θ is set to the median intensity value (=128 for 8-bit image), since value of a Haar-like feature becomes zero when μ_f is zero or maximum (=255). In practical object detection, feature values for positive samples are distributed in the range of low to medium intensities, while those for negative samples are distributed more widely up to high intensity ranges as shown in Fig. 2(c). As was introduced in Section 2.2, the probability that a feature is determined to 'negative' increases as feature value becomes smaller. Utilizing the above observation, false detections can be reduced without sacrificing true detections by decreasing feature values only for highly illuminated regions. This can be achieved by adjusting θ. If $\theta > 128$, the normalization factor becomes larger only at medium to high illumination.

3.3. Computational complexity and memory requirements

There is hardly any difference on computational complexity between (2) and (4). However, variance computation requires an additional integral image whose pixel values are sum of squared pixels of the original image. To generate an $m \times n$ sized integral image, at least 2 mn additions are required [15]. Pixel size of an integral image is at least 4 bytes for most sizes of images, and 4 bytes are not enough for the integral image of squared pixels, if image size is larger than 256×256. Table 1 shows computation and memory required for generating integral images. For an $m \times n$ input image, the proposed feature saves 2 mn additions and mn multiplications of computation and 8 mn bytes of memory by not using the integral image of squared pixels.

4. Experimental results

For experiments, we trained face classifiers, pedestrian classifiers for night vision, and vehicle classifiers employing the conventional Haar-like feature [2] without variance normalization (HF) and with variance normalization (NHF), the local binary pattern descriptor (LBP), and the proposed Haar Contrast Feature (HCF). Local binary pattern is well-known due to its robustness under

Table 1
Computation requirement and memory usage to generate integral images for an $m \times n$ input image.

	Computation requirement		Memory usage (bytes)
	# additions	# multiplications	
Haar-like feature with variance normalization	4 mn	mn	12 mn
Haar Contrast Feature	2 mn	–	4 mn

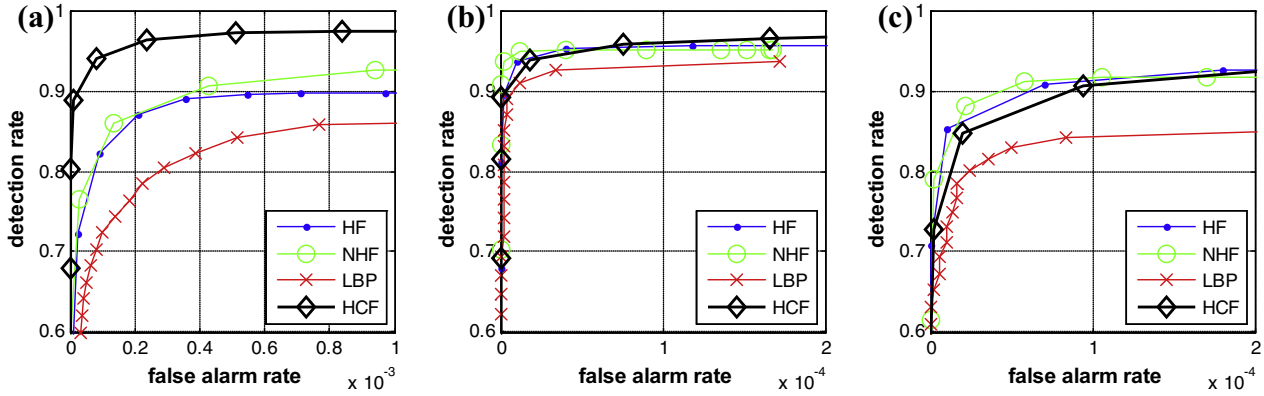

Fig. 4. ROC curves (a) face classifiers, (b) pedestrian classifiers, (c) vehicle classifiers.

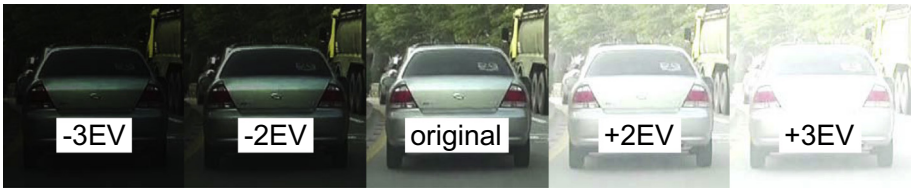

Fig. 5. Sample images whose exposure values were adjusted by −3 EV, −2 EV, +2 EV, and +3 EV.

illumination changes [29]. In our experiments, multi-scale block local binary pattern is used [30]. Face samples were obtained by cropping the images in Extended Yale Database B [21]. The Extended Yale Database B contains 16,128 images of 28 human subjects under 9 poses and 64 illumination conditions. All of the face samples were resized to 18×22. For pedestrian and vehicle, 15,000 and 25,000 samples were obtained by cropping the images captured at night time with near-infrared illumination and those captured on road with a vehicle mounted camera, respectively. All of the pedestrian and vehicle samples were resized to 14×28 and 20×20, respectively. For each of the classifiers, 5000 samples were used for training, and remaining samples were used for

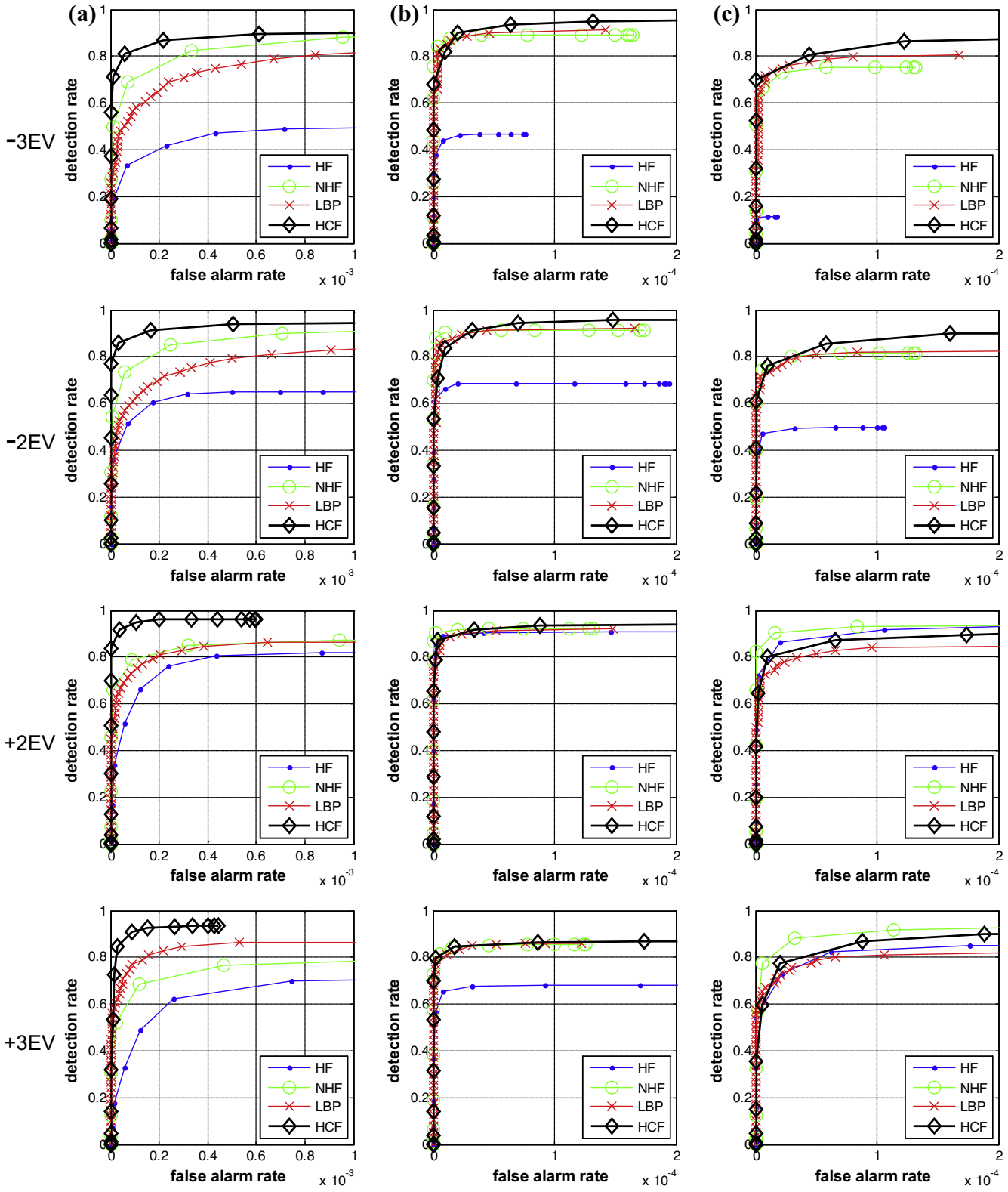

Fig. 6. Performance of classifiers under a wide range of illumination conditions. (a) Face classifiers, (b) pedestrian classifiers, (c) vehicle classifiers.

evaluating detection accuracy. Each classifier was trained using AdaBoost. For classifiers employing the proposed HCF, we trained them with several different thresholds ($\theta = 128$, 144, and 160), and selected classifiers which showed the best performance. Face classifier trained with $\theta = 160$, pedestrian classifier and vehicle classifier trained with $\theta = 144$ were used for the experiments.

4.1. Detection accuracy of classifiers

Fig. 4 shows receiver operating characteristic (ROC) curves. Face classifier employing the proposed HCF outperforms those employing HF, NHF or LBP in Fig. 4(a). Pedestrian classifier and vehicle classifier employing HCF show comparable performance to those employing HF or NHF, and outperform those employing LBP in Fig. 4(b) and (c). Images from the Extended Yale Database B were captured under various illumination conditions, while pedestrian and vehicle images were captured under relatively uniform illumination conditions. There exist strong shades in parts of the face for many samples. Experimental results confirm that the proposed feature descriptor shows improved performance under various illumination conditions.

4.2. Detection accuracy under various illumination conditions

For comparing performance of classifiers under a wide range of illumination conditions, we prepared additional samples by adjusting exposure values of each sample with a photo editing tool. Exposure values were adjusted by −3 EV, −2 EV, +2 EV, and +3EV. Fig. 5 shows sample images whose exposure values were adjusted.

Fig. 6 shows ROC curves obtained by experiments with the exposure-adjusted samples. Classifiers employing HF show much degraded performance for most samples, especially under −3 EV and −2 EV conditions. This result shows that Haar-like features cannot be used without correction (normalization) under varying illuminations. Those employing NHF show robust performance for the overexposed samples, but show degraded performance for the underexposed samples. On the contrary, classifiers employing the proposed HCF show robust performance for most samples. Experimental results show that performance of classifiers employing LBP is hardly degraded under most illumination conditions. However, classifiers employing HCF always outperform those employing LBP.

5. Conclusion

Haar-like feature has been popularly employed for object detection due to its simplicity and run-time efficiency. However, it has a fundamental limitation in that it is not illumination-invariant. In this paper, we propose an improved feature descriptor, *Haar Contrast Feature*, which can be used in replace of Haar-like feature for efficient object detection under various illumination conditions. The proposed feature computes contrast using the normalization factor devised to reflect the average intensity of feature region, and can also be computed fast using an integral image. For experiments, classifiers for face, pedestrian, and vehicle were trained by employing the conventional Haar-like feature, the local binary pattern descriptor, and the proposed feature descriptor, and their performances were evaluated. Experimental results confirm that the proposed feature descriptor shows improved performance under a wide range of illumination conditions, thus can be efficiently used for object detection in real-time environment.

References

[1] C. Papageorgiou, M. Oren, T. Poggio, A general framework for object detection, in: Proceedings of International Conference on Computer Vision, 1998, pp. 555–562.
[2] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 511–518.
[3] Y. Freund, R. Schapire, A short introduction to boosting, J. Jpn. Soc. Artif. Intell. 14 (5) (1999) 771–780.
[4] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, Ann. Stat. 28 (2) (1998) 337–407.
[5] T. Mita, T. Kaneko, B. Stenger, O. Hori, Discriminative feature co-occurrence selection for object detection, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1257–1269.
[6] I. Landesa-Vázquez, J. Alba-Castro, The role of polarity in Haar-like features for face detection, in: Proceedings of International Conference on Pattern Recognition, 2010, pp. 412–415.
[7] S. Chen, X. Ma, S. Zhang, AdaBoost face detection based on Haar-like intensity features and multi-threshold features, in: Proceedings of International Conference on Multimedia Signal Processing, vol. 1, 2011, pp. 251–255.
[8] P. Viola, M. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, Int. J. Comput. Vis. 63 (2) (2005) 153–161.
[9] X. Cui, Y. Liu, S. Shan, X. Chen, W. Gao, 3D Haar-like features for pedestrian detection, in: Proceedings of IEEE International Conference on Multimedia Expo, 2007, pp. 1263–1266.
[10] Y. Li, W. Lu, S. Wang, X. Ding, Local Haar-like features in edge maps for pedestrian detection, in: Proceedings of International Congress Image and Signal Processing, vol. 3, 2011, pp. 1424–1427.
[11] Y. Xin, S. Xiaosen, S. Li, A combined pedestrian detection method based on Haar-like features and HOG features, in: Proceedings of International Works Intelligent Systems and Applications, 2011, pp. 1–4.
[12] V. Hoang, A. Vavilin, K. Jo, Pedestrian detection approach based on modified Haar-like features and AdaBoost, in: Proceedings of International Conference on Control, Automation and Systems, 2012, pp. 614–618.
[13] P. Negri, X. Clady, S. Hanif, L. Prevost, A Cascade of boosted generative and discriminative classifiers for vehicle detection, EURASIP J. Adv. Signal Process. 2008 (136) (2008).
[14] A. Haselhoff, A. Kummert, A vehicle detection system based on Haar and triangle features, in: Proceedings of IEEE Intelligent Vehicles Symposium, 2009, pp. 261–266.
[15] J. Cui, F. Liu, Z. Li, Z. Jia, Vehicle localisation using a single camera, in: Proceedings of IEEE Intelligent Vehicles Symposium, 2010, pp. 871–876.
[16] M. Enzweiler, D. Gavrilla, Monocular pedestrian detection: survey and experiments, IEEE Trans. Pattern Anal. Mach. Intell. 31 (12) (2009) 2179–2195.
[17] P. Dollár, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: an evaluation of the state of the art, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 743–761.
[18] C. Wojek, B. Shiele, A performance evaluation of single and multi-feature people detection, in: Proceedings of DAGM Symposium, 2008, pp. 82–91.
[19] P. Dollár, Z. Tu, P. Perona, S. Belongie, Integral channel features, in: Proceedings of the British Machine Vision Conference, 2009, pp. 91.1–91.11.
[20] P. Belhumeur, D. Kriegman, What is the set of images of an object under all possible illumination conditions, Int. J. Comput. Vis. 28 (3) (1998) 245–260.
[21] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[22] J. Lee, B. Moghaddam, H. Pfister, R. Machiraju, A bilinear illumination model for robust face recognition, in: Proceedings of IEEE International Conference on Computer Vision, vol. 2, 2005, pp. 1177–1184.
[23] A. Cordiner, P. Ogunbona, W. Li, Illumination invariant face detection using classifier fusion, in: Proceedings of Pacific Rim Conference on Multimedia, 2008, pp. 456–465.
[24] X. Xie, W. Zheng, J. Lai, P. Yuen, Face illumination normalization on large and small scale features, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[25] T. Zhang, Y. Tang, B. Fang, Z. Shang, X. Liu, Face recognition under varying illumination using Gradientfaces, IEEE Trans. Image Process. 18 (1) (2009) 2599–2606.
[26] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, T. Poggio, Pedestrian detection using wavelet templates, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1997, pp. 193–199.
[27] B. Kisacanin, Integral image optimizations for embedded vision applications, in: Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation, 2008, pp. 181–184.
[28] P. Viola, M. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2) (2004) 137–154.
[29] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[30] S. Liao, X. Zhu, Z. Lei, L. Zhang, S. Li, Learning multi-scale block local binary patterns for face recognition, in: Proceedings of International Conference on Advances in Biometrics, 2007, pp. 828–837.