

Global Haar-Like Features: A New Extension of Classic Haar Features for Efficient Face Detection in Noisy Images

Mahdi Rezaei¹, Hossein Ziaei Nafchi², and Sandino Morales³

¹ The University of Auckland, New Zealand
m.rezaei@auckland.ac.nz

² Synchronmedia Laboratory, École de Technologie Supérieure, Canada
hossein.zi@synchronmedia.ca

³ The University of Auckland, New Zealand
sandinomorales@gmail.com

Abstract. This paper addresses the problem of detecting human faces in noisy images. We propose a method that includes a denoising preprocessing step, and a new face detection approach based on a novel extension of Haar-like features. Preprocessing of the input images is focused on the removal of different types of noise while preserving the phase data. For the face detection process, we introduce the concept of *global* and *dynamic global* Haar-like features, which are complementary to the well known classical Haar-like features. Matching dynamic global Haar-like features is faster than that of the traditional approach. Also, it does not increase the computational burden in the learning process. Experimental results obtained using images from the MIT-CMU dataset are promising in terms of detection rate and the false alarm rate in comparison with other competing algorithms.

Keywords: Face detection, Global Haar-like features, Phase-preserving denoising, AdaBoost.

1 Introduction

Face detection is the key step in many face analysis systems [21,23]. Current research is aiming at increasing the robustness of the detectors [22,16,9]. Among the proposed face detection algorithms, boosting-based detection with efficient use of integral image, Haar-like features and a cascade of weak classifiers, have defined high-performance systems [22,16,10].

Following the well-known Viola-Jones face detector [19], many researches have achieved further improvements in the performance of this detector. Currently, research in the field can be categorized into four subject areas:

1. Speeding up the learning process.
2. Speeding up the face detection.



Fig. 1. Detection results using the standard Haar-like detector. *Left:* Detection results on a “noisy” input image. *Right:* Detection results on the denoised image.

3. Defining a better trade-off between detection rate and false-positive rate.
4. Combining the three mentioned criteria.

For example, heuristic methods trying to improve the detection speed [13], or different version of AdaBoost like. Float boost [5], ChainBoost [20], cost-sensitive boosting [9,1], KLBoost [7], FCBoost [16], or RCECBoost [15] that aim at speeding up the AdaBoost convergence, or at improving the final performance of the detector.

Some variations of the Haar-like features have been proposed to improve the performance of boosting-based detectors [6,11,12]. These types of Haar-like features algorithms were introduced to deal with rotated faces and to improve the detection/false-positive rate. Any of these face detectors uses similar variations of the image preprocessing step suggested by Viola and Jones [19]. In that paper the authors utilized a fast variance-normalization preprocessing for face/non-face windows for dealing with illumination artifacts (as defined in [18] and [17]).

In this paper, we propose an image preprocessing step different to variance-normalization, to overcome the problem of noisy images and that of images contaminated with illumination artifacts. We use the image denoising method suggested by Kovese in [2]. This method is able to preserve the important *phase information* of the images, based on the non-orthogonal and complex-valued log-Gabor wavelets. Figure 1 shows an example of the results obtained with a standard Haar-like detector for a “noisy” input image (left), and on the denoised version of the same image (right).

We apply this technique of denoising in conjunction with a novel version of the Haar-like features method which together lead to an outperforming result. As the main contribution of the paper, we propose *global* Haar-like features which complement the commonly used standard Haar-like features. With global Haar-like features we introduce a new point of view to take benefit from the intensity information of the whole sliding query window. This is in contrast to the standard Haar-like features that only looks through dark-bright adjacent regions. We also propose *dynamic global* Haar-like features aiming to update global feature values based on intensity variations over the query image.

Since adding new features can increase the computational burden, only selected classical Haar-like features (during the learning process) are candidates for

becoming global features. We designed a face detector system through boosting denoised features with an efficient use of *standard* and *global* Haar-like features. Outstanding experimental results, in terms of the detection rate and the false-positive rate, show the robustness of the proposed method when using ideal, noisy, and illumination artifacts affected images.

The rest of the paper is organized as follows. Section 2 provides a discussion of the image denoising method based on a phase-preserving algorithm. Global and dynamic global Haar-like features are detailed in Section 3. Section 4 focuses on training a cascade of classifiers based on global and dynamic global features. Section 5 deals with experimental results, and Section 6 concludes.

2 Phase-Preserving Denoising of Images

A phase-preserving denoising method was proposed by Kovessi in [2]. It assumes that phase information of images is the most important feature and tries to preserve this information, of course by trying to keep the magnitude information, as well.

Let M_ρ^e and M_ρ^o denote the even-symmetric and odd-symmetric wavelets at a scale ρ which are known as quadratic pairs. Considering the responses from each quadrature pair of the filters, a resultant response vector is defined as follows:

$$[e_\rho(x), o_\rho(x)] = [f(x) * M_\rho^e, f(x) * M_\rho^o] \quad (1)$$

where $*$ denotes convolution, and values $e_\rho(x)$ and $o_\rho(x)$ are the real and imaginary parts in the complex-valued frequency domain. The amplitude of the transform at a given wavelet scale is given by:

$$A_\rho(x) = \sqrt{e_\rho(x)^2 + o_\rho(x)^2} \quad (2)$$

and the local phase is given by:

$$\varphi_\rho(x) = \tan^{-1} \left[\frac{o_\rho(x)}{e_\rho(x)} \right] \quad (3)$$

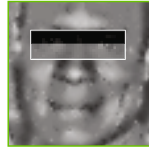
Having one response vector for each filter scale, there will be an array of such vectors for each pixel x in a signal. The denoising process includes defining an appropriate noise threshold for each scale as well as reducing the magnitudes of the response vectors, while maintaining the phase without any changes. The most important step of the denoising process is determining the thresholds. For this end, Kovessi [2] used the expected response of the filters to a pure noise signal.

If the signal is purely Gaussian white noise, then the position of the resulting response vectors from a wavelet quadratic pair of filters at some scale will form a 2D Gaussian distribution in the complex plane. Kovessi [2] showed that the distribution of the magnitude responses can be modelled by the Rayleigh distribution

$$R(x) = \frac{x}{\sigma_g^2} \exp \frac{-x^2}{2\sigma_g^2} . \quad (4)$$



$$F_{haar} = F_{white} - F_{black} = 15573 - 14733 = 840$$



$$F_{haar} = F_{white} - F_{black} = 23429 - 19326 = 4103$$



Haar feature applied in eye region

Fig. 2. Improved result for F_{haar} (integral image) after phase-preserving denoising. *Top:* Original image. *Middle:* Denoised image. *Bottom:* Sample of an applied Haar-like feature.

Also, the amplitude response from the smallest scale of the filter pair across the whole image will be the noise with Rayleigh distribution.

Finally by estimating the mean value μ_r and standard deviation σ_r of the Rayleigh distribution, the shrinkage threshold can be estimated. The thresholds are automatically determined and applied for each filter scale.

A number of parameters impacts the quality of the denoised output image. The threshold of noise standard deviations to be rejected (k), the number of filter scales to be used (N_ρ) and the number of orientations (N_r) are the key parameters.

We set the parameters $k = 3$, $N_\rho = 4$ and $N_r = 4$ in our experiments. These parameters result in an acceptable representation of small and middle-size faces. However, for large faces, it can lead to erroneous results. One approach is using a set of different parameters to obtain different images. Another approach is scaling the original images and then using the same parameters for conversions.

We used the second approach for a better speed-up. After a conversion to the denoised form, adaptive histogram equalization is used for both training and test images.

Figure 2 shows the discriminate advantage of using denoised images. The sample Haar-like feature applied on the eye region shows increased feature values (F_{haar}) which leads to a faster convergence of the classifier.

3 Global Haar-Like Features

Viola and Jones used five types of Haar-like features, which, from now on, we identify as *local features*. The value of a local feature f_i is computed by subtraction of white and black regions, $w_i - b_i$, using integral image [19].

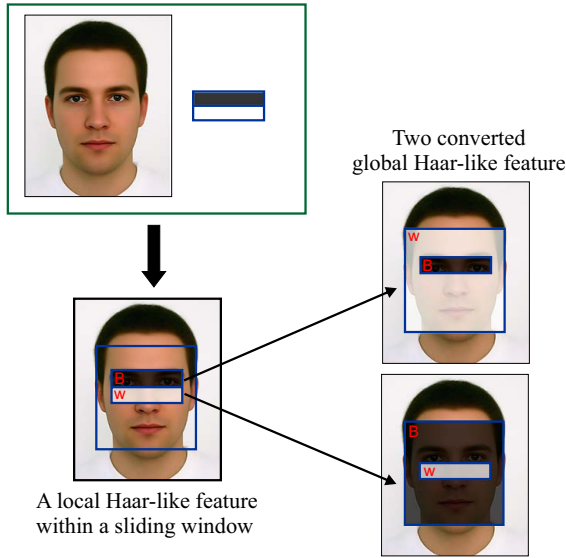


Fig. 3. Extraction of Global Haar-features from a standard Haar-feature

3.1 Global Features

For every given local feature, we introduce two global Haar-like features (in short, global features) as $F_b = F - b_i$ and $F_w = F - w_i$ where F is the integral value of the whole reference model (window). Global features will be used in conjunction with the local ones (Figure 3). We call them global features, as these feature values provides global information in addition to a standard (local) one.

If a local feature is being selected by a boosting algorithm for the formation of a cascade, then it would be a candidate to be a global feature as well. Global features are faster than local features, since the required values for calculating them are already computed at earlier stages. Figure 3 illustrates conversion steps from a local feature to global ones.

A problem with weak classifiers is that in the last stages, many of them are needed to reject 50% of non-face samples. The inclusion of these weak classifiers highly increases the computational cost. Therefore, global features are an efficient alternative, as they are faster to calculate and also because they provide a better classification, due to adding a new level of information by extracting different patterns than the local features.

In short, the term *global* in this paper refers to a comparison between the whole window and a portion of that window, while the common local features refer to *adjacent* rectangles of *equal* size.

3.2 Dynamic Global Features

Let b_i and w_i denote the integral values for the black and white regions of the i^{th} local feature, respectively, defining the local feature value as $w_i - b_i$. Let F

be the integral value of the reference model (sliding window). The current local feature is now accompanied by two global feature values, to be used in a weak classifier of the cascade for a given sliding window. In the dynamic version of global features, we update F by

$$F = F + \sum_{i=1}^{j \leq n} (w_i - b_i) \quad (5)$$

where, n is the total number of local features in the current cascade and j is current index over the global feature being assessed.

By using this equation, as the input windows progress through the cascade, the value of F is updated using the global features. We call this type of features *dynamic global Haar-like features* (in short, dynamic global features). Experimental results show that the dynamic global features can obtain a higher detection rate and less false positive rate in comparison with the non-dynamic version of the global features.

4 Boosting Cascades with Local and Global Features

In this section, a cascade of weak classifiers is designed by considering the global features. It is common that each stage of a cascade should reject 50% of negative samples while the true detection rate remains close to optimal.

When global features are considered, it is important to decide which of the local features should be considered as being global. One approach is to temporarily keep a current global feature and continue searching for the next local feature, without considering the effect of the current global feature. If global features show a better rejection rate, then it is efficient to choose the reserved global feature as the desired feature and then searching for next local features again. Also, even if their rejection rate becomes equal or near to equal, the global features are preferred.

Pseudocode for the learning cascades is provided as Algorithm 1. Applying the learning process, the following weak classifiers are obtained, where the optional pairs (ϕ_b^k, ϕ_w^k) denote global features:

$$(\theta_l^k, (\phi_b^k, \phi_w^k)), \dots, (\theta_l^n, (\phi_b^n, \phi_w^n)). \quad (6)$$

We observed that when not using dynamic global features, the number of global features selected during the cascade design is insignificant. Also, the effect of only using global features is not noticeable.

However, by using the dynamic global features, the number of global features selected was noticeable. The use of dynamic global features increases the performance of Haar-like feature based face detectors, in terms of detection rate, false alarm rate, and average number of features met in a window, and consequently results in a speed-up. Therefore, we preferred to use dynamic global features in conjunction with the local features.

Algorithm 1. Learning weak classifiers by using local and dynamic global features

Input: N_p positive samples; N_n negative samples.

Initialisation: Let $F_w = F_b = F$, where F is the sum of intensities in the whole window. Let $k = 1$.

Output: $(\theta_l^k, (\phi_b^k, \phi_w^k)), \dots, (\theta_l^n, (\phi_b^n, \phi_w^n))$.

1. Find k^{th} local weak classifier θ_l^k with threshold $T_l^k = \sum_{i=1}^{m_k} (w_i - b_i)$; where m_k is the total number of local features in the k^{th} classifier.
 2. Find next $(k + 1^{th})$ weak classifier θ_l^{k+1} ;
 3. Find k^{th} pair of global weak classifiers ϕ_b^k and ϕ_w^k , corresponding to the black and white parts of the local feature, respectively; set $T_b^k = \sum_{i=1}^{m_k} (F_b - b_i)$, and $T_w^k = \sum_{i=1}^{m_k} (F_w - w_i)$;
 4. Decide to choose best classifier(s) among (ϕ_b^k) , (ϕ_w^k) , and θ_l^{k+1} ;
 5. **if** A global classifiers is selected **then**
 6. Update the values of F_w and F_b as: $F_w = F_w + w_i$, $F_b = F_b - b_i$;
 7. Set $k = k + 1$, find the next local weak classifier θ_l^k ;
 8. Go to Step 3;
 9. **else**
 10. $k = k + 1$;
 11. Add θ_l^k to the cascade and search for next local weak classifier θ_l^{k+1} ;
 12. Go to Step 3;
 13. **end if**
-

5 Experimental Results

While a standard Haar-like classifier performs well in face detection under ideal lighting conditions [22], it seriously has difficulties when detecting faces under challenging illumination conditions, even for straight forward looking faces. However, we observed that after a phase-preserving denoising, and using the suggested global and dynamic global features, considerable improvements can be achieved in the true detection rate in comparison to that obtained using a standard Haar-like classifier. Choosing proper denoising parameters (k , N_ρ and N_r) can lead to further improvements.

In order to validate the methods proposed in this paper, we designed four classifiers using two preprocessing methods and different combinations of the local, global and dynamic global features.

The first detector was trained based on a variance-normalized samples using only the standard local Haar-like features (VN+Standard Local Haar).

The second classifier (VN+DyGlobal) was also learned from variance-normalized preprocessed samples, but this time using both local and dynamic global features.

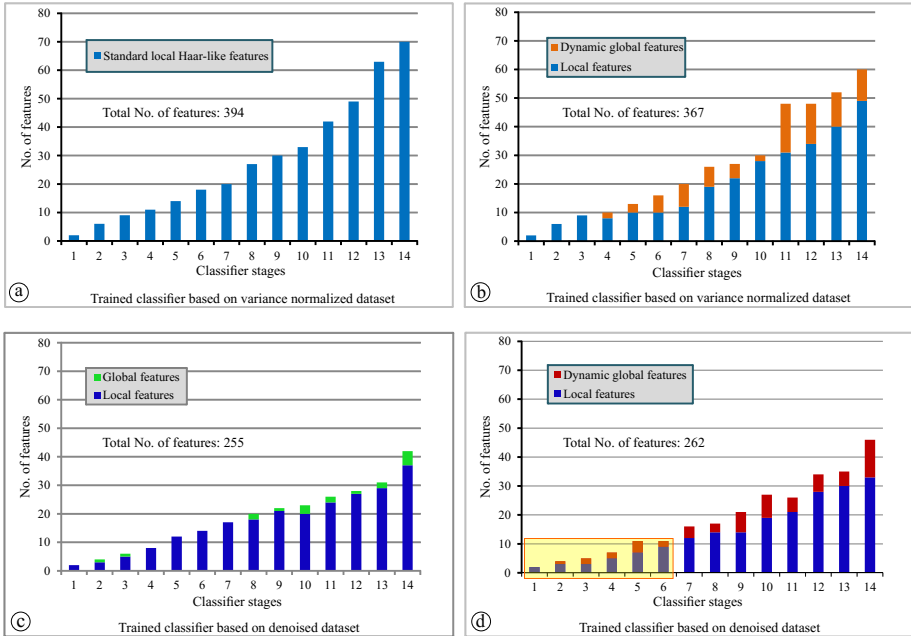


Fig. 4. Distribution of features for different learned cascades, each consists of 14 stages

The third detector is learned by local and global features, and the training dataset is enhanced by phase-preserving denoising technique (PPD+Global).

And the last detector (PPD+DyGlobal) uses local and dynamic global features, also based on the denoised training dataset.

For training all four classifiers, we used a large dataset of 10,000 face samples from different ages, genders, races and nationalities, mainly from AR [8] and Yale [4] datasets.

In addition, 50,000 non-face samples were considered as negative samples to train each stage of the cascade. Non-face samples were selected randomly from a dataset of scene categories [3]. This dataset consists of fifteen natural scene categories, including buildings, offices, roads, and landscapes.

We evaluated the performance of the four classifiers in terms of number of features involved in each stage, speed, and precision rate.

Figure 4 depicts the distribution graphs of local and global features for each of the four mentioned classifiers. Graph (a) is related to the first classifier (VN+Standard Local Haar) and it shows a total number of 394 features in a cascade of 14 weak classifiers. The graph shows that classifier (a) involves highest number of features, compared to other three classifiers. This means higher computational cost to confirm or reject a face candidate.

Graph (b) represents a faster classifier with a considerably smaller number of local features and also a total number of 367 features including both local and Dynamic global features, trained based on variance normalized dataset.

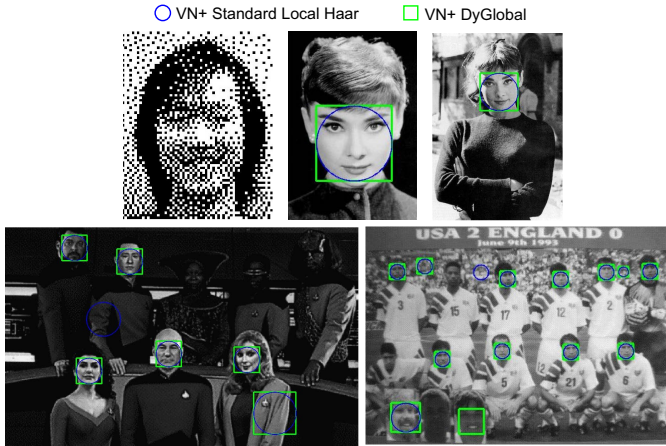


Fig. 5. Sample detection results for the first two classifiers trained based on standard Haar features (blue circles) or dynamic global features (green squares), using a variance-normalized dataset

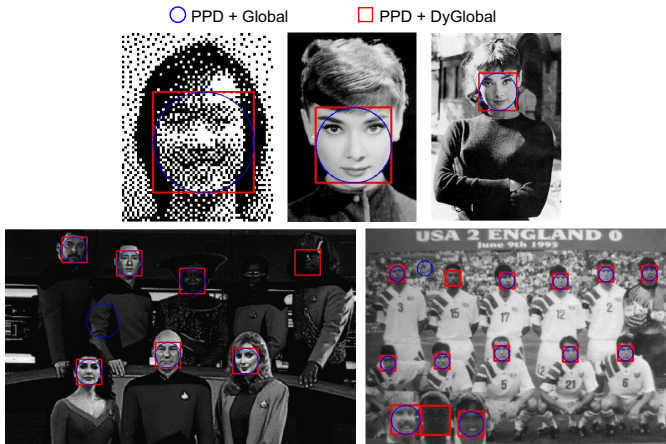


Fig. 6. Sample detection results for the last two classifiers trained based on global features (blue circles) or dynamic global features (red squares), using the proposed denoised dataset

Graphs (c) and (d) show feature distributions when we trained the classifiers with a phase-based denoised dataset. While the total number of features in graph (c) and (d) are very close to each other (255, and 262 respectively), the Classifier PPD+DyGlobal outperforms the other three classifiers as follows: Considering 50% rejection rate for each stage, 97% of non-face images will be rejected within the first six stages (weak classifiers), so having the minimum number of features in the first six stages plays a very crucial rule (i.e. the smaller the number of features, the faster the classifier). Classifier (d) contains only 40 features in its first six stages, while classifiers (a), (b), and (c) involve 60, 56, and 46 features in

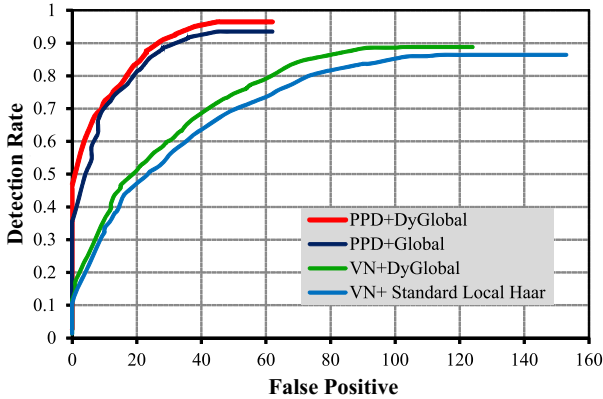


Fig. 7. ROC curve of the proposed detectors on MIT-CMU dataset. PPD+DyGlobal denotes the classifier trained based on denoised samples using the local and dynamic global features. PPD+Global denotes the classifier trained based on denoised samples using local and global features. Similarly, the two other detectors, are learned from variance normalized samples.

the six early stages which means 50%, 40%, and 15% slower operation compared to the PPD+DyGlobal classifier.

In addition to computational cost, we also need to consider effectiveness and accuracy of the four classifiers in terms of recall rate and false alarm.

Figures 5 and 6 show detection results for some sample images that were not used in the training phase in none of the four classifiers. The figures illustrate that the classifier we trained based on dynamic global classifier and phase-based denoising provides more accurate results. Figure 7 illustrates the *receiver operating characteristic* (ROC) curves for the four mentioned detectors evaluated on the MIT-CMU [24] dataset. The figure shows that the best results were obtained using the PPD+DyGlobal detector. Observe that PPD+Global also performed better than the considered local methods.

As further evaluation, Table 1 provides a comparison between the proposed PPD+DyGlobal classifier, standard Viola-Jones, as well as four other state-of-the-art face detection techniques [16,11,15,20]. The overall results confirm that the proposed method not only incorporates a smaller number of features (as discussed in the first part of the experimental section) but also outperforms the others in terms of a higher detection rate and a smaller number of false-positives due to global information provided by the proposed dynamic global features.

Table 1. Comparison between PPD+DyGlobal and state-of-the-art face detection techniques

Method	Proposed	VJ[19]	RCBoost[16]	Pham[11]	Opt[15]	Chain[20]
Detection rate (%)	96.4	86.3	94.00	91.00	95.00	93.00
# False-positives	62	153	96	170	900	130

6 Conclusions

In this research, we studied the effect of a preprocessing technique on input images for boosting-based face detectors. Phase-preserving denoising of images is used to preprocess the input images. The paper also proposed a new type of Haar-like features, called global Haar features, for the first time. This type of features supports faster calculations than local Haar features and provides a new level of global information from the query patches. We also proposed to use a dynamic version of the global features that updates its value dynamically as the classifier assess next local features.

Four distinct cascades are learned with and without using the denoised images and global Haar features. Finally, resulting detection rates and false-alarm rates shown a significant advantage for the proposed technique against state-of-the-art face detection systems, especially for challenging lighting conditions and noisy input images.

The proposed technique is expected to be effective on various object detection fields, as we also applied the proposed technique as part of a vehicle detection system [14] with superior results. We highly recommend to use, and to do further study on the proposed *global Haar-like features* for different kinds of boosting or Haar-based object detectors, not limited to face detection.

Acknowledgment. The authors thank professor Reinhard Klette for discussions and comments on the paper.

References

1. Hou, X., Liu, C.L., Tan, T.: Learning boosted asymmetric classifiers for object detection. In: Computer Vision and Pattern Recognition, pp. 330–338 (2006)
2. Kovese, P.: Phase Preserving Denoising of Images. Digital Image Computing, Techniques and Applications (1999)
3. Lazebnik, S., Schmid, C., Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. In: Computer Vision and Pattern Recognition, pp. 2169–2178 (2006)
4. Lee, K.C., Ho, J., Kreigman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Analysis Machine Intelligence 27, 684–698 (2005)
5. Li, S.Z., Zhang, Z.: Floatboost learning and statistical face detection. IEEE Trans. PAMI 26, 1112–1123 (2004)
6. Lienthart, R., Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection. In: International Conference on Image Processing, pp. 900–903 (2002)
7. Liu, C., Shum, H.Y.: Kullback-leibler boosting. In: Computer Vision and Pattern Recognition, pp. 587–594 (2003)
8. Martinez, A.M., Benavente, R.: The A.R. face dataset. CVC Technical Report (1998)
9. Masnadi Shirazi, H., Vasconcelos, N.: Cost-Sensitive Boosting. IEEE Trans. PAMI 33, 294–309 (2011)
10. Masnadi Shirazi, H., Vasconcelos, N.: High Detection-rate Cascades for Real-Time Object Detection. In: ICCV, pp. 1–6 (2007)

11. Pham, M.T., Cham, T.J.: Fast Training and Selection of Haar features using Statistics in Boosting-based Face Detection. In: International Conference on Computer Vision, pp. 1–7 (2007)
12. Pham, M.T., Gao, Y., Houg, V.T.D., Cham, T.J.: Fast Polygonal Integration and Its Application in Extending Haar-like Features to Improve Object Detection. In: Computer Vision and Pattern Recognition, pp. 942–949 (2010)
13. Rezaei, M., Klette, R.: Novel Adaptive Eye Detection and Tracking for Challenging Lighting Conditions. In: Park, J.-I., Kim, J. (eds.) ACCV Workshops 2012, Part II. LNCS, vol. 7729, pp. 427–440. Springer, Heidelberg (2013)
14. Rezaei, M., Terauchi, M.: Vehicle Detection Based on Multi-feature Clues and Dempster-Shafer Fusion Theory. In: 6th Pacific-Rim Symposium on Image and Video Technology (2013)
15. Saberian, M.J., Vasconcelos, N.: Learning Optimal Embedded Cascades. IEEE Trans. PAMI 34, 2005–2018 (2012)
16. Saberian, M.J., Vasconcelos, N.: Boosting Classifier Cascades. In: Neural Information Processing Systems (2010)
17. Struc, V., Vesnicer, B., Mihelic, F., Pavesic, N.: Removing illumination artifacts from face images using the nuisance attribute projection. In: IEEE International Conference on Acoustics Speech and Signal Processing, pp. 846–849 (2011)
18. Vaudrey, T., Morales, S., Wedel, A., Klette, R.: Generalized Residual Images Effect on Illumination Artifact Removal for Correspondence Algorithms. Pattern Recognition 44, 2034–2046 (2011)
19. Viola, P., Jones, M.: Robust Real-Time Face Detection. International Journal of Computer Vision 57, 137–154 (2004)
20. Xiao, R., Zhu, L., Zhang, H.J.: Boosting chain learning for object detection. In: Computer Vision and Pattern Recognition, pp. 709–715 (2003)
21. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE Trans. PAMI 24, 34–58 (2002)
22. Zhang, C., Zhang, Z.: Boosting-Based Face Detection and Adaptation. Morgan & Claypool (2010)
23. Zhang, C., Zhang, Z.: A Survey of Recent Advances in Face Detection. Microsoft Research. Technical Report, MSR-TR-2010-66 (2010)
24. Carnegie Mellon University image dataset,
http://www.vasc.ri.cmu.edu/idb/html/face/frontal_images