Multi-Pose Face Detection with Asymmetric Haar Features

Geovany A. Ramirez
University of Texas at El Paso
500 W University Ave - El Paso TX 79968

Olac Fuentes
University of Texas at El Paso
500 W University Ave - El Paso TX 79968

Abstract

In this paper we present a system for multi-pose face detection. Our system presents three main contributions. First, we introduce the use of asymmetric Haar features. Asymmetric Haar features provide a rich feature space, which allows to build classifiers that are accurate and much simpler than those obtained with other features. The second contribution is the use of a genetic algorithm to search efficiently in the extremely large parameter space of potential features. Using this genetic algorithm, we generate a feature set that allows to exploit the expressive advantage of asymmetric Haar features and is small enough to permit exhaustive evaluation. The third contribution is the application of a skin color-segmentation scheme to reduce the search space. Our system uses specialized detectors in different face poses that are built using AdaBoost and the C4.5 rule induction algorithm. Experimental results using the CMU profile test set and BioID frontal faces test set, in addition to our own multi-pose face test set, show that our system is competitive with other systems presented recently in the literature.

1. Introduction

Face detection is an important first step for applications in several areas, including biometrics, human-computer interfaces, and surveillance. Face detection is a difficult task, due to different factors such as varying size, orientation, poses, facial expression, occlusion and lighting conditions [24]. Most recent approaches to this problem pose it as a binary classification problem, where one needs to classify every window in an image as belonging to the class of interest (i.e. faces), or not. Recent research contributions have focused mostly on two main aspects of the problem: feature engineering and classifier design. Feature engineering consists of designing and selecting classes of features that can improve the performance of the classifiers that are based on them. Work on classifier design consists of adapting existing classification algorithms to the detection and recognition problems, or of designing special-purpose algorithms that are targeted to these problems.

In a series of papers, Viola and co-workers advocated an approach to object recognition based on Haar features, which are equivalent to the difference of the sum of the intensity levels of two contiguous equal-sized rectangular image regions. They presented an algorithm for computing these features in constant time, which makes them suitable for real-time performance. Using Haar features, a cascade of classifiers based on the Adaboost algorithm was constructed, yielding accurate classification, albeit at the expense of long training times. Successful applications of this methodology were presented in face detection [21], image retrieval [18], and pedestrian detection [22].

In this paper we propose three extensions to the work by Viola and co-workers. First, we introduce the use of asymmetric Haar features, eliminating the requirement of equal-sized positive and negative regions in a feature. We propose Haar features with asymmetric regions that can have regions with either different width or height, but not both. This results in a more expressive feature space, which, as we will show, allows to build classifiers that are much simpler than those obtained with the standard features. While the number of symmetric Haar features is large (around 180,000 in Viola's work), it is still feasible to perform exhaustive evaluation of these features in order to build a classifier. On the other hand, if we use asymmetric features, the number of potential features grows to over 200 million for a 24×24 pixel window, which makes exhaustive evaluation impossible. The second contribution of this work is the use of a genetic algorithm to search efficiently in the parameter space of potential features. Using this genetic algorithm, we generate a small feature set that allows to exploit the expressive advantage of asymmetric features that is small enough to permit exhaustive evaluation. The third contribution is the application of a skin color-segmentation scheme to reduce the search space. Using color segmentation, we can speed-up the classification process by eliminating from consideration all windows that do not contain regions that are similar to skin.

We present experimental results showing the application of our method to two test sets that have been used commonly in the literature, the CMU profile test set, and BioID frontal test set. The experiments show that our method can attain better results than other methods while generating significantly simpler classifiers.

2. Related Work

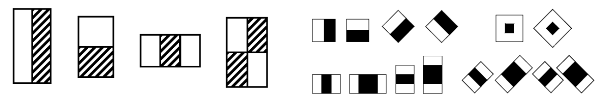
Rowley et al. developed a frontal face detection system that scanned every possible region and scale of an image using a window of 20×20 pixels [15]. Each window is pre-processed to correct for varying lighting; then a retinally connected neural network is used to process the pixel intensity levels of each window to determine if it contains a face. In later works, they provided invariance to rotation perpendicular to the image plane by means of another neural network that determined the rotation angle of a region and; then the region was rotated by the negative angle and given to the original neural network for classification [16].

Schneiderman and Kanade detected faces and cars from different view points using specialized detectors [17]. For faces, they used 3 specialized detectors for frontal, left profile, and right profile views. For cars, they used 8 specialized detectors. Each specialized detector is based on histograms that represent the wavelet coefficients and the position of the possible object, and then they used a statistical decision rule to eliminate false negatives.

Fröba and Ernst used the census transform as a feature for face detection [3]. The census transform builds a bit-vector that functions as a descriptor of the 3-by-3 neighborhood of each pixel. The bit value of a neighbor is one if the pixel has an intensity that is greater than the average intensity of the neighborhood; otherwise it is zero. Using these simple features and a cascade-style classifier, they obtained results that were comparable to the best systems presented to date.

Viola et al. used Haar features, as described above, in their object detection systems [21, 18, 22]. For classification, they used a modified version of the Adaboost algorithm, an ensemble method originally presented by Freund and Schapire [2], which has proved to yield excellent results in several learning domains. Adaboost builds a sequence of classifiers, where the ith classifier in the sequence is biased to correct the misclassifications of classifiers $1, ..., i-1$. The final classification of Adaboost is given by the weighted average of the classification made by each of the individual classifiers. The fact that classifiers that appear later in the sequence attempt to correct the mistakes made by earlier ones results in accuracy levels that are superior to most other algorithms, particularly in situations where noise levels are moderate.

Li and Zhang present another approach to detect frontal and profile faces using Haar features and the FloatBoost algorithm [9]. Their system can detect faces with ± 45 degree rotation perpendicular to the image plane using 10 detectors specialized for different face poses. In [10], a set of specialized support vector machines were trained to detect faces at specific angles in the view sphere, and, when an image needed to be classified, another support vector machine detected the pose and chose the appropriate specialized detector to use.



(a) (b)

Figure 1. (a) Haar features introduced by [21]. (b) Extension to the basic set proposed by [11]

Wu et al. detect frontal and profile faces with arbitrary in-plane rotation and up to 90-degree out-of-plane rotation [23]. They used Haar features and a look-up table to develop strong classifiers. To create a cascade of strong classifiers they used Real AdaBoost, an extension to the conventional AdaBoost . They built a specialized detector for each of 60 different face poses. To simplify the training process, they took advantage of the fact that Haar features can be efficiently rotated by 90 degrees or reversed, thus they only needed to train 8 detectors, while the other 52 can be obtained by rotating or inverting the Haar features.

Treptow and Zell present an object detection system based on Haar features and the Adaboost algorithm [19]. They add 2 more types of Haar features to the basic set of Viola and Jones and replace the exhaustive search for an evolutionary search in each iteration of Adaboost. They can speed up the training process for about 3 times using evolutionary search.

3. Haar Features

Haar features are based on Haar wavelets, which are functions that consist of a brief positive impulse followed of a brief negative impulse. In image processing, a Haar feature is the difference between the sum of all pixels in two o more regions. Papageorgiou et al. were the first to use Haar features for face detection [12]. They used three types of Haar features of size 2×2 and 4×4 pixels, for a total of 1,734 different features in a 19×19 face image. Viola and Jones proposed a basic set of four types of Haar features that are shown in Figure 1a [21]. The value of Haar feature is given by the sum of intensities of the pixels in the light region minus the sum of intensities in the dark region. Using all possible sizes, they generate around 180,000 features for a 24×24 pixel image. Lienhart and Maydt presented an extension to the basic set with rotated Haar features as we can see in Figure 1b [11].

Using a straightforward implementation, the time required to perform the sum of pixels increases linearly with the number of pixels. Viola and Jones proposed to use the

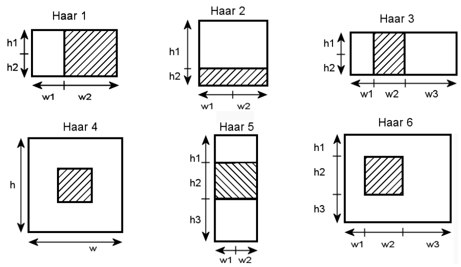


Figure 2. Assymetric Haar features used.

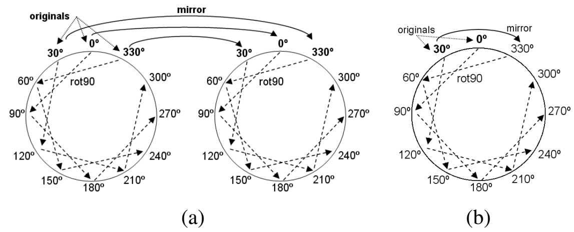


Figure 3. (a) Specialized detectors for profile faces. (b) Specialized detector for frontal faces.

integral image as preprocessing to compute the sum of any region of any size in constant time [21]. Each element of the integral image contains the sum of pixels in the original image that are above and to the left of that pixel; using this idea allows to compute a two-region Haar feature using only six memory access and a three-region feature Haar with only eight.

We propose an extension for basic Haar features, which we call *Asymmetric region Haar features*, which are shown in Figure 2. In contrast with basic Haar features, these new features can have regions with different width or height, but not both. It will be shown that these features are able to capture defining characteristics of objects more accurately than traditional ones, allowing the development of simpler and more effective classifiers. By allowing asymmetry, the number of possible configurations for Haar features grows exponentially and is an overcomplete set. For the 6 Haar features shown in Figure 2, there are around 200 million possible configurations for a 24×24 image. Using all the possible configurations is unfeasible, therefore, to deal with this limitation we propose to use a Genetic Algorithm to select a subset of features. Details will be presented in the next section.

4. Specialized Detectors

We use specialized detectors for faces in frontal, left profile and right profile poses. In addition, we use 12 specialized detectors, one every 30 degrees to cover the 360 degree in-plane rotation of each pose. Therefore, we have a total of 36 specialized detectors. Each specialized detector consists of a cascade of Strong Classifiers (SC) created with the AdaBoost algorithm used by Viola and Jones [21]. We use weak classifiers (WC) based on the C4.5 rule induction algorithm [13] that is associated with only one Haar feature. We use the same strategy used by Wu et al. to create all the specialized detectors [23], rotating 90 degree and inverting the Haar features of a few specialized detectors, as shown in Figure 3.

4.1. Selecting Haar with a Genetic Algorithm

Training a specialized detector using all the possible configurations of the asymmetric region Haar features would be impractical. For instance, using an initial training set of 3,000 examples including faces and non-faces and all the possible Haar configurations, we would take about 6 months using a 2GHz Pentium 4 computer for only one specialized detector. Therefore, to reduce the number of possible Haar features we use a Genetic Algorithm (GA) to select the width and the height of a feature. In the GA, one individual is a weak classifier (WC) that contains only one Haar feature and is trained with the C4.5 algorithm. An individual representing the Haar parameters is shown in Table 1. The fitness of each individual corresponds to the classification error on an initial training set. We compute a WC for each place on the image and for each type of feature, for a total of 2,431 Haar features (see Table 2). Since we can use a binary or decimal representation, we performed two experiments. In the first experiment we use a binary representation and a Genetic Algorithm with Elitism. If after a crossover, or mutation, a new individual showed an out of range configuration, the individual is penalized with a low fitness. In the second experiment we use a decimal representation that avoids the creation of invalid individuals by crossover. For mutation, we generate a uniformly distributed random value inside the allowed range. In both experiments we use two point crossover with a deterministic crossover model presented in [8]. This model consists of combining the best and the worst individuals in the population, then the second best with the second worst, and so on. With this crossover model, it is possible to perform a larger exploration on the search space. The average results of 24 tests for each experiment are shown in Table 3. For each experiment we use an initial population of 20 individuals and 20 generations. The best result was obtained using a decimal representation with a 10% mutation rate. Using the GA to select a subset of Haar features, we can reduce training time to 9 hours on our 2GHz Pentium 4 computer; this corresponds to a 99.8% reduction in time.

4.2. Training a Specialized Detector

A specialized detector is a cascade of Strong Classifiers (SC). We can create a cascade of SCs using a variation of the algorithm used by Wu et al. [23], where we introduce the GA to reduce the feature set. Our algorithm is presented in Figure 4. After obtaining a set of WCs with the GA, new examples are added to the training set and then the AdaBoost

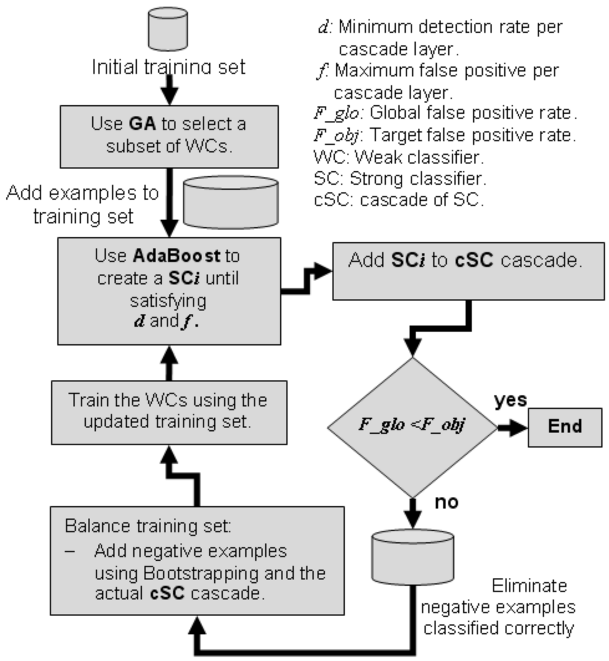


Figure 4. Algorithm to train a specialized detector.

Table 1. Individuals parameters for each type of Haar

	Parameters
Haar 1	$\{h_1, h_2, w_1, w_2\}$
Haar 2	$\{h_1, h_2, w_1, w_2\}$
Haar 3	$\{h_1, h_2, w_1, w_2, w_3\}$
Haar 4	$\{h_1, w_1\}$
Haar 5	$\{h_1, h_2, h_3, w_1, w_2\}$
Haar 6	$\{h_1, h_2, h_3, w_1, w_2, w_3\}$

Table 2. Number of WCs for each location in a image of 24×24 pixels.

Haar	1	2	3	4	5	6	Total
X **range**	[2, 20]	[1,21]	[3,19]	[3,19]	[1,21]	[2,21]	
Y **range**	[1,23]	[3,23]	[1,23]	[3,21]	[3,21]	[2,23]	
WCs	437	441	391	323	399	440	2,431

Table 3. Average results of 24 tests for the 2 experiments with different mutation percentages.

Mutation	Experiment 1 error	Experiment 2 error
10%	**0.1622**	**0.1570**
30%	0.1668	0.1581
50%	0.1682	0.1603
70%	0.1758	0.1575

algorithm is used to create a new SC. AdaBoost terminates when the minimum detection rate and the maximum false positive rate per layer in the cascade are attained. If the target false positive rate is achieved, the algorithm ends. Otherwise, all negative examples correctly classified are eliminated and the training set is balanced adding negative examples using a bootstrapping technique. With the updated training set, all the WCs are retrained and then used in AdaBoost.

5. Skin Color Segmentation

Skin color segmentation is a technique that has previously been used for face detection (for example, [7, 1, 20]), but due to the high accuracy required, these methods are slow. In our system, we are interested in using skin color segmentation only to reduce the search space in color images. Therefore, we require a fast method that provides a very low false negative rate, while high false positive rates may be acceptable. We use a skin color segmentation based on the *YCbCr* color space and one simple rule to classify every pixel as skin or non-skin. The rule was defined statistically using a set of 33 images that contain 81 persons of different skin tones in different lighting conditions. The skin color pixels were segmented manually and then used to define a range for channels Cb and Cr. A pixel is classified as belonging to skin if both its Cb and Cr components are within two standard deviations of the mean values for skin pixels found in the training images. This is illustrated in Equation 1, where $\bar{y}$ is the mean and σ is the standard deviation for a given channel. The resulting rule is shown in Equation 2.

$$\bar{y} - 2\sigma \leq \mathrm{C} \leq \bar{y} + 2\sigma \tag{1}$$

$$img_{x,y} = \begin{cases} 1, \quad if & (i) \quad\quad 80 \leq \mathrm{Cb_{x,y}} \leq 140 \\ & (ii) \quad\quad 135 \leq \mathrm{Cr_{x,y}} \leq 170 \\ 0, & otherwise \end{cases} \tag{2}$$

We tested our skin color segmentation method using 271 color images under different lighting conditions and with persons with different skin tones. The method can eliminate about 65% of all the pixels in the images, preserving the face region. The skin color segmentation method does not eliminate all the regions that do not contain skin, but it reduces the search space by 65% with only 4 comparisons per pixel.

6. Experimental Results

For frontal detectors we use the same training set to select a subset of Haar features and to train the specialized detectors. For profile detectors we use an initial training set to select a subset of Haar features and an extended training set to create the specialized detectors. For each training set, we performed a careful selection of images that represent, as much as possible, the variation of faces. We use faces of males and females, with different ages, of different races, with and without structural components such as glasses and beard, and different lighting conditions and sources. Also, we added non-face images by randomly selecting regions in images without faces. We normalized each set to 0 degrees of rotation in the plane and thus can generate a training set for any angle by only rotating the images. To increase the number of images, we use the mirror image of each image.

Table 4. Number of images used for frontal and profile detectors.

pose	original images	variations (initial/extended)	Total faces	non-faces	Total
Frontal	593	19	23,720	23,720	47,740
Profile	458	8/17	4,122/8,244	4,693/16,227	8,815/24,471

Table 5. Results for 137 images from CMU test set.

	Test 1. 2 specialized detector for left and right profile in $\pm 15°$.	**Test 2.** 6 specialized detector for left and right profile in $\pm 45°$.
Number of WC in each SC	0°:[16, 34, 162, 200, 200]	0°:[16, 34, 162, 200, 200] 30°:[27, 38, 164, 200, 200] 330°:[25, 38, 200, 200, 200]
Detection rate	**82.2%**	**92.5%**
False positives	148	432

Table 6. Comparison with others works on CMU profile test set.

	Detection rate	**False positives**
Schneiderman and Kanade[17]	92.8%	700
Wu et al. [23]	91.3%	415
Test 1	82.2%	148
Test 2	92.5%	432

Table 7. Results for BioID test set.

	Detection rate	**False positives**	**Number of Haar features per layer**
Test 1	80.47%	92	[17, 39, 200, 200, 200]
Test 2	83.89%	126	[17, 39, 200, 200]
Test 3	93.68%	432	[11, 18, 37, 300, 300]

Table 8. Comparison with others works that used the BioID test set.

	Detection rate	**False Positives**
Jesorsky et al. [5]	91.80%	Not reported
Kirchberg et al. [6]	92.80%	Not reported
Hamouz et al. [4]	91.30%	Not reported
Fröba and Ernst [3]	97.75%	25
Ramirez and Fuentes [14]	93.23%	2236
our system test 3	93.68%	432

Table 9. Comparison with others works.

	Number of stages in cascade	**Number of Haar features in all the cascade**
Viola and Jones [21]	32	4,297
Wu et al. [23]	16	756
our detector	5	666

Table 10. Multi-pose test set results.

	Without skin color segmentation	**With skin color segmentation**
Detection rate	91.93 %	95.03 %
False positives	110	40

Since each specialized detector will be trained to cover a range of 30 degrees, we generate variation of each image rotating it in the range of ± 15 degrees. In Table 4 we show the number of training images used for detectors in frontal and profile poses.

To test the profile detectors we used a subset from the CMU profile test set that consists of 137 images with 214 profile faces with a rotation angle of ± 30 degrees [17]. We did two tests; the first one was using two specialized detectors in left and right profile respectively and ± 15 degrees. In the second test, we used six specialized detectors to cover the left and right profile in the range of ± 45 degrees. The results are shown in Table 5. Using six specialized detectors we increase the detection rate but also the number of false positives. In Table 6, we compare our results with other works. To test the frontal detectors we use the BioID test set that consist of 1,521 grayscale images of 384×288 pixels, with a frontal view of 23 different persons under a high variety of lighting conditions, backgrounds and face sizes [5]. The results of multiple tests are shown in Table 7. In Table 8 we compare our results with the BioID test with other works.

In Table 9 we present a comparison between our specialized detector in frontal faces and the detectors presented in [21] and [23]. We can see in Table 9 that our detector uses fewer Haar features than the others. We can conclude that asymmetric Haar features are better than symmetric Haar features to represent face appearance. Additionally, the use of the genetic algorithm for feature selection enables us to use a much richer feature space without incurring increased computational costs.

To the best of our knowledge, a standard test set for multi-pose face detection is not currently in use. Therefore, we created our own multi-pose test set. Our test set consists of 45 color images with 60 profile faces and 101 frontal faces. The faces have a rotation of up to ± 45 degrees. We performed two tests, the first test without the skin color segmentation and the second with the skin color segmentation. In both tests we used nine specialized detectors to cover the rotation range of ± 45 degrees for faces in frontal and profile pose. In Table 10 we show the results of the 2 tests. We can see in Table 10, that by using the skin color segmentation we can reduce the number of false positives and obtain a higher detection rate.

In Figures 5, 6 and 7 we show some representative results for the CMU profile set, the BioID set, and our multi-pose set, respectively.

7. Conclusions and Future Work

In this paper we have presented a face detection system that introduces three extensions to previous state-of-the-art systems. First, we introduced asymmetric Haar features as an extension to the basic set of Haar features. According to our experiments, these new features can represent face appearance more accurately than previously used ones. Second, we reduced the training time using a genetic algorithm that allows to exploit the expressive advantage of asymmetric features. With the small feature set generated by the genetic algorithm we can deal with large training sets in 0.02% of the time that would be required using the full feature set. Third, using our skin color segmentation we can reduce the search space, resulting in faster processing and fewer false positives. Our system can detect faces in different poses

Figure 5. Some results from the CMU profile test set.

Figure 6. Some results from the BioID test set.

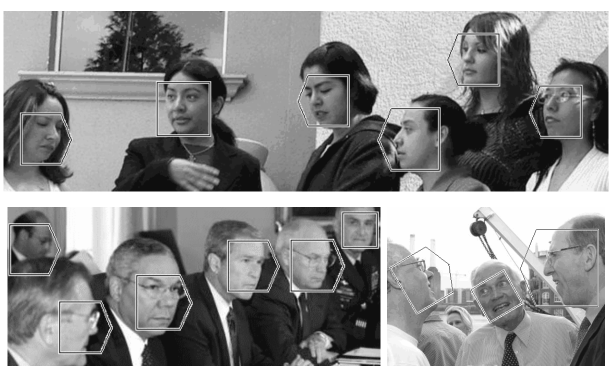


Figure 7. Some results from our multi-pose test set.

with a detection rate of up to 92.5% for profile faces and up to 93.68% for frontal faces.

Future work will be oriented to test our system in other object detection problems such as car detection and street detection in aerial images. In addition, we will perform experiments using other boosting algorithms such as FloatBoost and Real AdaBoost. We will also use more types of Haar features and other optimization algorithms such as particle swarm optimization.

References

[1] D. Chai and A. Bouzerdoum. A bayesian approach to skin color classification in ycbcr color space. In *TENCON*, volume 2, pages 421–424, 2000.

[2] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, pages 148–156, 1996.

[3] B. Froba and A. Ernst. Face detection with the modified census transform. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 91– 96, Erlangen, Germany, May 2004.

[4] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, and H. Kälviäinen. Affine-invariant face detection and localization using gmm-based feature detector and enhanced apparence model. In *6th International Conference on Automatic Face and Gesture Recognition*, pages 67–72, 2004.

[5] O. Jesorsky, K. Kirchberg, and R. W. Frischholz. Robust face detection using the hausdorff distance. In *Third International Conference on Audio- and Video-based Biometric Person Authentication*, Lecture Notes in Computer Science, pages 90–95. Springer, 2001.

[6] K. J. Kirchberg, O. Jesorsky, and R. W. Frischholz. Genectic model optimization for hausdorff distance-based face localization. In *International Workshop on Biometric Authentication*, pages 103–111. Springer, 2002.

[7] J. Kovac, P. Peer, and F. Solina. Illumination independent color-based face detection. In *3rd International Symposium on Image and Signal Processing and Analysis*, volume 1, pages 510–515, 2003.

[8] A. F. Kuri-Morales. Efficient compression from non-ergodic sources with genetic algorithms. In *Fourth Mexican International Conference on Computer Science*, pages 324–329, 2003.

[9] S. Z. Li and Z. Zhang. Floatboost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1112–1123, 2004.

[10] Y.-M. Li, S.-G. Gong, J. Sherrah, and H. Liddell. Support vector machine based multi-view face detection and recognition. *Image and Vision Computing*, 22(5):413–427, May 2004.

[11] R. Lienhart and J. Maydt. An extended set of Haar-like features for rapid object detection. In *International Conference on Image Processing*, volume 1, pages I–900–903, 2002.

[12] C. P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, page 555, Washington, DC, USA, 1998. IEEE Computer Society.

[13] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1993.

[14] G. A. Ramirez and O. Fuentes. Face detection using combinations of classifiers. In *The 2nd Canadian Conference on Computer and Robot Vision, 2005*, pages 610–615, 2005.

[15] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[16] H. A. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *Proceedings of 1998 IEEE Conference on Computer Vision and Pattern Recognition*, pages 38–44, Santa Barbara, CA, June 1998.

[17] H. Schneiderman and T. Kanade. A statistical model for 3-d object detection applied to faces and cars. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2000.

[18] K. Tieu and P. Viola. Boosting image retrieval. *International Journal of Computer Vision*, 56(1-2):17–36, 2004.

[19] A. Treptow and A. Zell. Combining adaboost learning and evolutionary search to select features for real-time object detection. In *IEEE Congress on Evolutionary Computation 2004*, volume 2, page 21072113, 2004.

[20] M.-J. S. D. Valaparla and V. Asari. Neural network based skin color model for face detection. In *32nd Applied Imagery Pattern Recognition Workshop*, pages 141–145, 2003.

[21] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of 2001 IEEE International Conference on Computer Vision and Pattern Recognition*, pages 511–518, 2001.

[22] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, 63(2):153–161, 2005.

[23] B. Wu, H. Ai, C. Huang, and S. Lao. Fast rotation invariant multi-view face detection based on real adaboost. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 79–84, 2004.

[24] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis ans Machine Intelligence*, 24(1):34–58, 2002.