

# Real-Time Face Detection Using Illumination Invariant Features

Klaus Kollreider, Hartwig Fronthaler, and Josef Bigun

Halmstad University, Box 823, SE-30118, Halmstad, Sweden  
{klaus.kollreider, hartwig.fronthaler, josef.bigun}@ide.hh.se  
<http://www.hh.se>

**Abstract.** A robust object/face detection technique processing every frame in real-time (video-rate) is presented. A methodological novelty are the suggested quantized angle features (“quangles”), being designed for illumination invariance without the need for pre-processing, e.g. histogram equalization. This is achieved by using both the gradient direction and the double angle direction (the structure tensor angle), and by ignoring the magnitude of the gradient. Boosting techniques are applied in a quantized feature space. Separable filtering and the use of lookup tables favor the detection speed. Furthermore, the gradient may then be reused for other tasks as well. A side effect is that the training of effective cascaded classifiers is feasible in very short time, less than 1 hour for data sets of order  $10^4$ . We present favorable results on face detection, for several public databases (e.g. 93% Detection Rate at  $1 \times 10^{-6}$  False Positive Rate on the CMU-MIT frontal face test set).

**Keywords:** Object detection, Face Detection, Biometrics, Direction Field, Orientation Tensor, Quantized Angles, Quangles, AdaBoost.

## 1 Introduction

When attempting to *detect* faces (or *locate* a single face) in a visual representation, image-based and landmark-based methods may be primarily distinguished between [1,2]. This paper focuses on the detection of frontal faces in 2D images and is assigned to the former category. Features here represent measurements made by means of some basis functions in a multidimensional space which should be contrasted to the term “facial features” sometimes used in the published studies to name subparts of a face, e.g. the eyes, mouth, etc., which we refer to as “landmarks”. Challenges in face detection are generally comprised of varying illumination, expression changes, (partial) occlusion, pose extremities [1] and requirements on real-time computations.

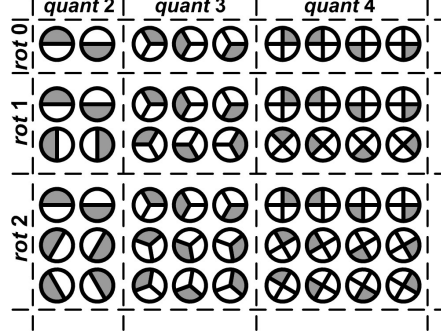
The main characteristics of still image-based methods is that they process faces in a holistic manner. Faces are learned by training on roughly aligned portraits as well as non-face-like images, and no parts of the face are intentionally favored to be used for face detection. The specific statistical pattern recognition method employed characterizes published studies. A popular approach uses

the so-called Eigenfaces [3], or the PCA (Principal Component Analysis) coordinates, to quantify the “faceness” of an image (region). More recent face detection systems employed neural networks [4, 5], or Support Vector Machines (SVM) as [6] in [7] to classify image regions as face or non-face. Also, a naive Bayes scheme was implemented in [8], and recently in [9], whereas an AdaBoost procedure [10] was concurrently adapted in [11, 12]. In principle, these techniques are not specific to detect faces in an image, but can be trained in an analogous manner to detect other objects, e.g. cars. The AdaBoost based face detection in [11, 12] has been suggested as being real-time, and has been followed up with other studies extending it to multi-poses, and reducing classifier complexity, e.g. [13, 14]. However, the employed features play a decisive role besides the used classifiers. Of all the published methods, only a few use the gray values directly for classification, but rather features. However, almost all approaches use a preprocessing of the gray values (e.g. histogram equalization or normalization) to minimize the effect of adverse light conditions, at the expense of computational processing. The methods suggested by [11, 13, 14] use Haar-like rectangle features, translating into a high detection speed whereas [12, 9] employed edge features with arguably lower execution speed. The recent method of [15] proposed binary coded local histograms (LBF) as features. A novelty in this study is the use of gradient angles only driven by the observation, that the gradient angle as opposed to the magnitude is, simply put, naturally robust to illumination changes. Gray value preprocessing becomes redundant, and we extend the illumination resilience by two contributions: First, the use of hierarchical and adaptive quantization levels improves the detection performance. Second, we do not only exploit the gradient angle, but also the structure tensor direction [16], encoding local orientation. Because we use quantized angle features, we term the latter “quangles” for expediency. Furthermore, these quangles are boosted in layers of a decision cascade as in [11], enabling also small classifiers. We achieve scale invariance through signal theoretically correct downsampling in a pyramidal scheme. The usefulness of our technique is shown in the context of face detection. A methodological advantage of the suggested scheme is the readily availability of some filtered signals for differential algorithms, for example, optical flow estimation, exploited for immediate person “liveness” [17] assessment. In comparison, the rectangle features suggested in [11], despite their value in pure object detection in still images, have limited reusability when it comes to other tasks. For a survey of landmark-based methods, which focus on a few salient parts, landmarks, e.g. the single eyes, mouth, nose of the face, we refer to [18, 1]. We present experimental results on several public databases, namely the MIT-CMU [5] and the YALE [19] face test sets.

## 2 Object Detection

### 2.1 The Quantized Angle Features (Quangles)

In this section we present the features for object detection, which we call “quangles”, representing quantized angle features. The gradient of an image is given



**Fig. 1.** Example of a set of quangle masks (angle displayed in polar form). The gray shaded areas correspond to the partitions yielding value 1 in equation (2).

in equation 1,

$$\nabla f = \begin{pmatrix} f_x \\ f_y \end{pmatrix} \tag{1}$$

where  $f_x$  and  $f_y$  denote the derivatives in x and y direction respectively. Furthermore,  $|\nabla f|$  indicates the magnitude of the gradient and  $\angle \nabla f$  refers to its angle. For the sake of object detection, we disregard the magnitude or intensity since it is highly affected by undesired external influences like illumination variations. The key instrument of our quangle features are the *quangle masks*, which are denoted as follows:

$$Q(\tau_1, \tau_2, \phi) = \begin{cases} 1, & \text{if } \tau_1 < \phi < \tau_2, \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The thresholds  $\tau_1$  and  $\tau_2$  constitute the boundaries of a partition in  $[0, 2\pi]$ . The quangle mask yields 1 if an angle  $\phi$ , is located within such a partition and 0 otherwise. In order to produce a set of quangle masks, we divide the full angle range  $[0, 2\pi]$  into an increasing number of quantizations (partitions), which are additionally rotated. An example is depicted in figure 1. A set of quangle masks  $\{Q\}_{maxQuant, numRot}$  is fully determined by the maximum number of quantizations  $maxQuant$  and rotations  $numRot$ . The parameter  $maxQuant$  has to be interpreted cumulatively, meaning that all quangle masks with less quantization steps are included in the set as well. The second parameter,  $numRot$ , indicates the number of rotations included in addition to each basic quangle mask. For example the final row in figure 1 corresponds to  $\{Q\}_{4,2}$ , which consists of 27 different quangle masks. In order to create such a set of quangle masks the thresholds  $\tau_1$  and  $\tau_2$  of each partition need to be determined. This can be done in a three step procedure:

1. First we define a sequence of threshold pairs  $\alpha_1$  and  $\alpha_2$  delimiting the desired number of partitions  $nQuant$  in the interval  $[0, 2\pi]$ , disregarding the

rotational component, by  $\alpha_1 = \frac{2\pi}{nQuant} \cdot quant$  and  $\alpha_2 = \frac{2\pi}{nQuant} \cdot (quant + 1)$ , where  $quant \in \{0, \dots, nQuant - 1\}$ .

2. In the second step we create the final threshold sequence for  $nQuant$  containing the threshold pairs  $\tau_1$  and  $\tau_2$ . For each partition  $quant$  we include the number of rotations up to  $numRot$ , by  $\tau_k = \text{mod} \left( \alpha_k - \frac{\pi}{nQuant} \cdot rot, 2\pi \right)$ , where  $rot \in \{0, \dots, numRot\}$  and  $k \in \{1, 2\}$ .
3. Performing the first two steps corresponds to creating a single cell in figure 1. In order to produce a complete quangle set, the two steps above need to be repeated for  $nQuant = \{2, \dots, maxQuant\}$ .

To detect objects in a single scale we use a sliding window approach, where an image is scanned by a so-called search or detection window. In order to look for candidates, the quangle masks need to be assigned to positions  $(i, j)$  within the detection window  $x$ . This defines at the same time our *quangle features*. We furthermore distinguish between two different types: Equation (3a) describes a quangle feature using the original gradient angle, whereas in equation (3b) double angle representation is employed.

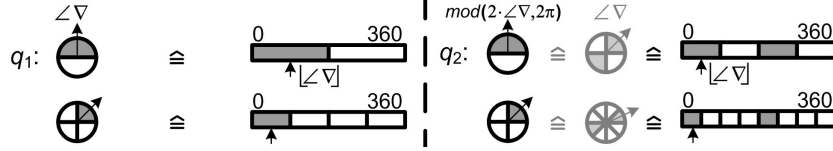
$$q_1(x, i, j, \tau_1, \tau_2) = Q(\tau_1, \tau_2, \angle \nabla x(i, j)) \quad (3a)$$

$$q_2(x, i, j, \tau_1, \tau_2) = Q(\tau_1, \tau_2, \text{mod}(2 \cdot \angle \nabla x(i, j), 2\pi)) \quad (3b)$$

Both quangle feature types in the equations above take the detection window  $x$ , the position  $(i, j)$  within  $x$  and a particular quangle mask out of  $\{Q\}$ . Using both,  $q_1$  and  $q_2$ , the number of possible features is determined by the size of the detection window and the number of employed quangle masks. We include both, single and double angle representation in our set of quangle features since they are meaningful at different sites within the search window. The original gradient is more informative within the object, e.g. between landmarks of a face, because it distinguishes between dark-light and light-dark transitions. The double angle representation maps  $\phi$  to  $2\phi$ , and has been shown to represent the structure tensor eigenvector directions [16]. Thereby  $\nabla f$  and  $-\nabla f$  are equivalent and represent orientations of linear structures, e.g. lines. The double angle representation is more resistant to illumination changes, especially helpful at object boundaries (background changes). Accordingly, both single angle and double angle features are complementary and meaningful features to represent objects.

## 2.2 Classifier Building

A good classification (yielding low error rate) cannot be obtained with a single quangle feature, but obviously, it is neither meaningful nor practical to evaluate all of them within the detection window. In order to find the most suitable quangles we employ AdaBoost [10]. In the process, a number of good features (termed weak classifiers) are combined, yielding a so-called strong classifier. Following the discrete AdaBoost algorithm, we select the weak classifier



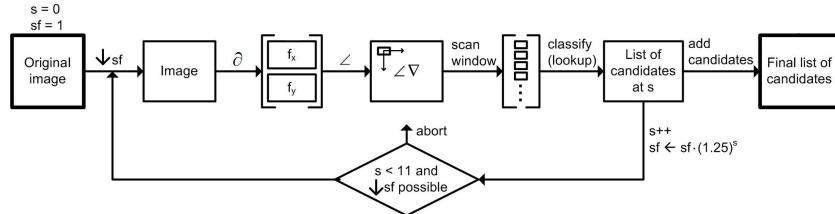
**Fig. 2.** Example lookup tables (stripes) representing quangle masks in case of single angle (left hand side) and double angle (right hand side) representation

$h_t(x) = q_{k_t}(x, i_t, j_t, \tau_{1_t}, \tau_{2_t})$  which minimizes the error  $e_t = \min_{i,j,\tau_1,\tau_2,k} \sum_l w_l \cdot |q_k(x, i, j, \tau_1, \tau_2) - y_l|$  over the training set (indexed by  $l$ ) in round  $t$  of the feature selection ( $y$  denotes the true class “1” or “0” and  $w$  weights the training examples). Eventually, we obtain a strong classifier which is composed of  $T$  weak classifiers, and each of the latter has a say in the final decision depending on the individual error  $\alpha_t = \log \frac{1 - e_t}{e_t}$ . While generally improving the detection/false positive rates, adding more and more weak classifiers unfortunately directly affects the classification time. An alternative to a single strong classifier is the so-called cascaded classifier scheme, a series of less complex strong classifiers, which is computationally efficient [11]. A single negative decision at any level of such a cascade leads to an immediate disregard of the concerned candidate image region. When training a strong classifier and adding it to the cascade, we apply a bootstrapping strategy. Previously rejected negative class examples are replaced by new ones, which the current cascade would (wrongly) classify as positive examples.

Another important factor for the training of such a cascade is time. Viola&Jones, for example, reported that the training time of their final classifier was in the order of weeks on a single machine. This was due to the large amount of rectangle features, necessary there, in combination with finding an eligible threshold for each of them. Employing our features, the training of a comparable cascade takes about an hour on an ordinary desktop computer, because less features suffice (quangles build upon derivative features) and the expensive calibration is skipped.

### 2.3 Implementation

Cascaded classifiers favor processing time in that only a few strong classifications accrue per image site. Furthermore, we can reduce the number of operations needed to calculate and classify a single feature. In this study, we employ so-called lookup tables to speed up this process. Recalling the quangle features of type  $q_1$  and particularly  $q_2$  in equations (3a) and (3b), lookup tables provide an effective solution for both of them. Figure 2 depicts two exemplary lookup tables for both  $q_1$  and  $q_2$ . Each quangle mask is represented by a binary lookup table, generated off-line. Gray shaded areas correspond to 1 and are defined by the respective quangle mask. The original gradient angle is used as table index, therefore we floor it to integer values in  $[0, 360[$ . However, the quangle features of



**Fig. 3.** The proposed object detection process

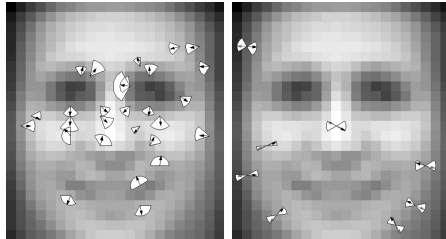
type  $q_2$ , need some further attention. As visualized,  $\text{mod}(2 \cdot \angle \nabla, 2\pi)$  corresponds to  $\angle \nabla$ , by means of “helper quangles” (displayed alleviated), only existing in the form of lookup tables. To construct one such, we half the thresholds of the original quangle mask. The resulting partition together with a  $180^\circ$  shifted version define the lookup table for  $q_2$ . As a consequence, 1 array access is needed per weak classification for any quangle.

The whole detection process is illustrated in figure 3. The image to be analyzed serves as a starting point at scale  $s = 0$  and scale factor  $sf = 1$ . We approximate the gradient (see equation (1)) of the whole image using separable Gaussians and their derivatives and extract the angle information. After this, we scan the image with the detection window to be classified using a trained cascade and the lookup tables introduced above. Having the candidates of the first scale, we successively reduce the image size by a factor of 1.25 and start over with the gradient calculation and window scanning, repeating like this for 10 times. The candidates from each scale are integrated. In order to eliminate multiple detection, neighboring candidates in position and scale are grouped.

## 2.4 Face Detection

In this section we apply the object detection system introduced in sections 2.1 and 2.2 to face detection. The size of the search window for face detection is  $22 \times 24$ . Our system operates in real-time at a resolution of  $640 \times 480$  using 11 scales on a standard desktop computer. We have been collecting approximately 2000 faces of varying quality from online newspapers for training purposes. All face images were aligned and artificially rotated in the interval  $[-10^\circ, \dots, +10^\circ]$  for pose resilience. Some background is included in a typical positive (face) example. On the other hand, the negative examples are chosen randomly from a large amount of images, which do not contain any faces.

In order to strengthen the argument in section 2.1, where we suggest the use of both single and double angle features, we train a strong classifier employing both feature types. This will also help us to pre-confine the feature space, in an attempt to prevent overtraining and to support feature selection a priori. Empirical tests on a subset of positive examples (900) and 9000 negative examples revealed that  $\{Q\}_{8,5}$  is an eligible set of quangle masks for face detection. The least number of quangle features to separate this 9900 examples errorfree

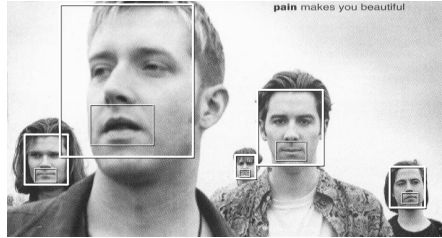


**Fig. 4.** A strong classifier employing 36 both, single and double angle features, which are displayed (projected lookup-tables) side by side

served as a criterion, besides economic parameters for the set  $\{Q\}$ . By doing so, we also advanced to reduce the complexity of strong classifiers. A number of 36 quangle features (28 of type  $q_1$  and 8 of type  $q_2$ ) were selected in the case of using  $\{Q\}_{8,5}$ . Figure 4 visualizes the selected features in separate detection windows. In both cases the black arrows indicate  $\angle \nabla$  of the underlying average training face. The white partitions show the range, the respective angle is supposed to be in. The hourglass-shaped partitions in the second image indicate double angle features, where the gradient could also have pointed in the opposite direction (gray arrows). The radii are modulated by  $\alpha_t$ , the weight of the corresponding weak classifiers. It can be observed that single angle features frequently occur in the inner facial regions, whereas features of the second type are situated in the bounding regions. In a further step, we trained two strong classifiers using the same training setup, yet employing either features of type  $q_1$  or  $q_2$ . Error-free separation of the training data involved 49 single angle or 83 double angle features, thus clearly favoring the combined setup. Other studies have suggested schemes for reducing classifier complexity [13, 14], which we did not investigate yet, because our features resulted in small classifiers. In a related study, [9], using a naive Bayes classifier, the single gradient angle was quantized into 7 partitions without a further study of flexible and lower quantization levels. In [12], no quantization but integer conversion was done and only the doubled gradient angle was used. Furthermore, the weak classifiers were different there, involving significantly more operations. We show an example of face (and mouth) detection by our method in figure 5.

### 3 Experiments

For the experiments, the face detector was configured as follows: The size chosen for the detection window was  $22 \times 24$  and the employed quangle masks were in  $\{Q\}_{8,5}$ . A cascaded classifier, comprising 22 levels, was trained on 2000 faces and 4000 non-faces (refilled). The total number of weak classifiers in the cascade was 700. Such a classifier complexity is very small compared to a couple of thousands as reported in [11, 20]. In operation, the first two levels of the cascade, comprising only 3 and 5 quangle features, respectively, are already able to reject 75% of all



**Fig. 5.** A (cropped) example image from the CMU-MIT face test set, with faces and mouths detected by the proposed method



**Fig. 6.** Two images from the YALE face test set, illustrating "severe" illumination changes, managed by the proposed method though

non-faces. Furthermore, we used  $s = 0$  (original resolution) as the starting scale and  $sf = 1.1$  as the factor for downsizing.

The performance of the face detection system detailed above is benchmarked on two publicly available databases, namely the YALE [19] and the CMU-MIT [5] face test sets. Extreme illumination and expression changes are the main challenge of the former test set, which consists of 165 frontal face images of 15 subjects. Table 1 shows the detection rates and the number of false positives of our method together with the ones for the face detection algorithm proposed in [9], on the YALE face test set. The results on the YALE test set confirm

**Table 1.** Detection rates and the number of false positives on the YALE face test set

| Method          | Detection Rate | False Positives |
|-----------------|----------------|-----------------|
| Nguyen [9]      | 86,6%          | 0               |
| Proposed method | 100%           | 0               |

that our face detection method is resistant to substantial illumination changes without performing any (histogram related) preprocessing. Note, that the latter is actually done in all methods we compare our results to. In figure 6, two "YALE faces" are shown, with indicated detections by the proposed method. Note the severity of the illumination conditions.

The CMU-MIT frontal face test set is among the most commonly used data sets for performance assessment of face detection systems. It is composed of 130 images, containing 507 frontal faces in total. The quality of the images, as well



as the scale of faces (compare figure 5) vary substantially here. In addition to the detection rate, this set also permits to give representative numbers for the false positives, because of many high resolution images. In table 2, the results of our technique on the CMU-MIT frontal test set are related to those of two prominent face detectors [11,5], by adjusting the false positive rate to a common level. Also, the detection rate achieved by our method for 1 false detection per million evaluated windows is given, constituting our best result.

**Table 2.** Detection and false positive rates on the CMU-MIT frontal face test set

| Method           | Detection Rate | False Positive Rate   |
|------------------|----------------|-----------------------|
| Rowley [5]       | 89,2%          | $1,27 \times 10^{-6}$ |
| Viola&Jones [11] | 92,9%          | $1,27 \times 10^{-6}$ |
| Proposed method  | 94,2%          | $1,25 \times 10^{-6}$ |
| Proposed method  | 93%            | $1 \times 10^{-6}$    |

## 4 Conclusion

In this study, we presented a novel real-time method for face detection. However, the technique is possible to be used as a general image-object detector, as current experiments indicate. The introduced quantized angle (“quangle”) features were studied experimentally and we presented evidence for their richness of information measured by their discriminative properties and their resilience to the impacts of severe illumination changes. They need no preprocessing, e.g. histogram equalization, histogram normalization, adding to their computational advantage. This is achieved by considering both the gradient direction and orientation, yet ignoring the magnitude. A quantization scheme is presented to reduce the feature space prior to boosting, i.e. it enables fast evaluation (1 array access). Scale invariance was implemented through an image pyramid. The training excels in rapidness, which enables the use of our object detector for changing environments and application needs. The practicability of the proposed methods and ideas was corroborated by satisfying experimental results for face detection (e.g. 93% Detection Rate at  $1 \times 10^{-6}$  False Positive Rate on the CMU-MIT frontal face test set).

## References

1. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting Faces in Images: A Survey. *IEEE-PAMI* 24(1), 34–58 (2002)
2. Hamouz, M., Kittler, J., Kamarainen, J., Paalanen, P., Kalviainen, H., Matas, J.: Feature-Based Affine-Invariant Localization of Faces. *PAMI* 27(9), 1490–1495 (2005)
3. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognitive Neuroscience* (Winter) 3(1), 71–86 (1991)
4. Sung, K.K., Poggio, T.: Example Based Learning for View-Based Human Face Detection. Technical report, Cambridge, MA, USA (1994)

5. Rowley, H., Baluja, S., Kanade, T.: Human Face Detection in Visual Scenes. In: *Advances in Neural Information Processing Systems 8*. pp. 875 – 881 (1996)
6. Cortes, C., Vapnik, V.: Support-Vector Networks. *Machine Learning* 20, 273–297 (1995)
7. Osuna, E., Freund, R., Girosi, F.: Training Support Vector Machines: an Application to Face Detection. In: *IEEE Conference on Computer Vision and Pattern Recognition* (1997)
8. Schneiderman, H., Kanade, T.: Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition. *cvpr 00*,45 (1998)
9. Nguyen, D., Halupka, D., Aarabi, P., Sheikholeslami, A.: Real-Time Face Detection and Lip Feature Extraction Using Field-Programmable Gate Arrays. *SMCB* 36(4), 902–912 (2006)
10. Freund, Y., Shapire, R.E.: A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences* 5(1), 119–139 (1997)
11. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features (2001)
12. Froeba, B., Kueblbeck, C.: Real-Time Face Detection Using Edge-Oriented Matching. In: Bigun, J., Smeraldi, F. (eds.) *AVBPA 2001*. LNCS, vol. 2091, pp. 78–83. Springer, Heidelberg (2001)
13. Li, S.Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H., Shum, H.: Statistical Learning of Multi-view Face Detection. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2350, pp. 67–81. Springer, Heidelberg (2002)
14. Sochman, J., Matas, J.: WaldBoost Learning for Time Constrained Sequential Detection. In: *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 150–156. IEEE Computer Society Press, Washington, DC, USA (2005)
15. Hadid, A., Pietikäinen, M., Ahonen, T.: A Discriminative Feature Space for Detecting and Recognizing Faces. In: *CVPR (2)*. pp. 797–804 (2004)
16. Bigun, J.: *Vision with Direction*. Springer, Heidelberg (2005)
17. Kollreider, K., Fronthaler, H., Bigun, J.: Evaluating Liveness by Face Images and the Structure Tensor. In: *Fourth IEEE Workshop on Automatic Identification Advanced Technologies AutoID 2005*, Buffalo, New York, pp. 75–80 (2005)
18. Smeraldi, F., Bigun, J.: Retinal vision applied to facial features detection and face authentication. *Pattern Recognition Letters* 23, 463–475 (2002)
19. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *PAMI* 19(7), 711–720 (1997)
20. Li, J., Wang, Y., Tan, T., Jain, A.K.: Live face detection based on the analysis of Fourier spectra. In: *Biometric Technology for Human Identification*, SPIE vol. 5404, pp. 296–303 (2004)