A Look at Eye Detection for Unconstrained Environments

Brian C. Heflin, Walter J. Scheirer, Anderson Rocha and Terrance E. Boult

Key words: Unconstrained Face Recognition, Eye Detection, Machine Learning, Correlation Filters, Photo-head Testing Protocol

1 Introduction

Eye detection is a necessary processing step for many face recognition algorithms. For some of these algorithms, the eye coordinates are required for proper geometric normalization before recognition. For others, the eyes serve as reference points to locate other significant features on the face, such as the nose and mouth. The eyes, containing significant discriminative information, can even be used by themselves as features for recognition. Eye detection is a well studied problem for the *constrained face recognition problem*, where we find controlled distances, lighting, and limited pose variation. A far more difficult scenario for eye detection is the *unconstrained face recognition problem*, where we do not have any control over the environment or the subject. In this chapter, we will take a look at eye detection for the latter, which encompasses problems of flexible authentication, surveillance, and intelligence collection.

A multitude of problems affect the acquisition of face imagery in unconstrained environments, with major problems related to lighting, distance, motion and pose. Existing work on lighting [1, 2] has focused on algorithmic issues (specifically, normalization), and not the direct impact of acquisition. Under difficult acquisition

Brian C. Heflin, Walter J. Scheirer, and Terrance E. Boult
Vision and Security Technology Lab, University of Colorado at Colorado Springs, Colorado 80918, USA, e-mail: lastname@uccs.edu

Anderson Rocha
Institute of Computing, University of Campinas (Unicamp), Campinas, Brazil, e-mail: anderson.rocha@ic.unicamp.br

Pre-print of chapter to appear in the book *Pattern Recognition, Machine Intelligence and Biometrics*. The original publication is available at http://www.springerlink.com.

circumstances, normalization is not enough to produce the best possible recognition results - considerations must be made for image intensification, thermal imagery and electron multiplication. Long distances between the subject and acquisition system present a host of problems, including high f-numbers from very long focal lengths, which significantly reduces the amount of light reaching the sensor, and a smaller amount of pixels on the faces, as a function of distance and sensor resolution. Further, the interplay between motion blur and optics exasperates the lighting problems, as we require faster shutter speeds to compensate for the subjects movement during exposure, which again limits the amount of light reaching the sensor. In general, we'll have to face some level of motion blur in order for the sensor to collect enough light. Pose variation, as is well known, impacts the nature of facial features required for recognition, inducing partial occlusion and orientation variation, which might differ significantly from what a feature detector expects.

Both lighting and distance should influence sensor choice, where non-standard technologies can mitigate some of the problem discussed above. For instance, EMCCD sensors have emerged as an attractive solution for low-light surveillance (where low-light is both conditional, and induced by long-range optics), because they preserve a great bit of detail on the face and can use traditional imagery for the gallery (as opposed to midwave-IR sensors). This makes them very attractive for biometric application as well. However, the noise induced by the cooling of the sensor also presents new challenges for facial feature detection and recognition. In this chapter, for the reasons cited above, we use the EMCCD to acquire our difficult imagery under a variety of different conditions, and apply several different eye detectors on the acquired images.

In Section 2 we take a brief survey of the existing literature related to difficult detection and recognition problems, as well as the pattern recognition works relevant to the detection techniques discussed in this chapter. In Section 3 we introduce a machine learning based approach to feature detection for difficult scenarios, with background on the learning and feature approach used. In Section 4 we introduce the correlation filter approach for feature detection, including a new adaptive variant. Our experimental protocol is defined in Section 5, followed by a thorough series of experiments to evaluate the detection approaches. Finally, in Section 6, we make some concluding remarks on our examination of algorithms for difficult feature detection.

2 RELATED WORK

On the algorithm front, we find only a few references directly related to difficult facial feature detection and recognition. Super-resolution and deblurring were considered in [3] as techniques to enhance images degraded by long distance acquisition (50m - 300m). That work goes further to show recognition performance improvement for images processed with those techniques compared to the original images. The test data set for outdoor conditions is taken as sequential video under daylight

conditions; the super-resolution process considers direct sequences of detected faces from the captured frames. The problem with this approach is that under truly difficult conditions, as opposed to the very controlled settings of [3] (full frontal imagery, with a constant inter-ocular distance), it is likely that a collection of detected faces in a direct temporal sequence will not be possible, thus reducing the potential of such algorithms. Real-time techniques to recover facial images degraded by motion and atmospheric blur were explored in [4]. The experiments of [4] with standard data sets and live data captured at 100m showed how even moderate amounts of motion and atmospheric blur can effectively cripple a facial recognition system. The work of [5] and [4] is more along the lines of what is explored in this paper, including a thorough discussion of the underlying issues that impact algorithm design, as well as an explanation of how to perform realistic controlled experiments under difficult conditions, and algorithmic issues such as predicting when a recognition algorithm is failing in order to enhance recognition performance.

In the more general pattern recognition literature, we do find several learning techniques applied to standard data sets for eye detection. Many different learning techniques have been shown to be quite effective for the eye detection problem. The work of [6] is most closely related to the learning technique presented in this work in a feature sense, with PCA features derived from the eyes used as input to a neural network learning system. Using a data set of 240 images of 40 different full frontal faces, this technique is shown to be as accurate as several other popular eye detection algorithms. [7] uses color information and wavelet features together with a new efficient Support Vector Machine (eSVM) to locate eyes. The eSVM, based on the idea of minimizing the maximum margin of misclassified samples, is defined on fewer support vectors than the standard SVM, which can achieve faster detection speed and comparable or ever higher detection accuracy [7]. The method of [7] consists of two steps. In the first step selects possible eye candidate regions using a color distribution analysis in YcbCr color space. The second validation step consists of applying 2D Haar wavelets to the image for multi-scale image representations followed by PCA for dimensionality reduction and using the eSVM to detect the center of the eye. [8] uses normalized eye images projected onto weighted eigenspace terrain features as features for an SVM learning system. [9] uses a recursive non-parametric discriminant feature as input to an AdaBoost learning system.

For recognition, a very large volume of work exists for correlation,but we find some important work on feature detection as well. Correlation filters [10, 11] are a family of approaches that are tolerant to variations in pose and expression, making them quite attractive for detection and recognition problems. Excellent face recognition results have been reported for the PIE data set [12] and the FRGC data set [13]. For the specific problem of eye detection, [14] first demonstrated the feasibility of correlation filters, while [15] introduced a more sophisticated class of filters that are more insensitive to over-fitting during training, more flexible towards training data selection, and more robust to structured backgrounds. All of these approaches have been tested on standard well-known data sets, and not the more difficult imagery we consider in this chapter. We discuss correlation in detail in Section 4.

Of course we should reduce the impact of difficult conditions using better sensors and optics, which is why we choose to use EMCCD sensors to allow faster shutter speeds. For the optics, one possibility gaining attention is the use of advanced Adaptive Optics (AO) models [16], which have proved effective for astronomy, though most do not apply to biometric systems. Astronomy has natural and easily added artificial "point sources" for guiding the AO process. Secondly, astronomical imaging is vertical, which changes the type and spatial character of distortions. More significantly, they have near point sources for guides, allowing for specialized algorithms for estimation of the distortions. Horizontal terrestrial atmospheric turbulence is much larger and spatially more complex making it much more difficult to address. To date, no published papers discuss an effective AO system for outdoor biometrics. While companies such as AOptix[1] have made interesting claims, public demonstrations to date have been stationary targets indoor at less than 20m, where there is no atmospherics and minimal motion blur .

A critical limiting question for adaptive optics is the assumption of wave-front distortion and measurement. For visible and NIR light, the isoplanatic angle is about 2 arc seconds (0.00027 degrees or motion of about 0.08mm at 50m). Motion outside the isoplanatic angle violates the wave-front model needed for AO correction [17]. An AO system may be able to compensate for micro-motion on stationary targets, where a wave-front isoplanatic compensation AO correction approach would be a first-order isoplanatic approximation to small motions, but it's unclear how it could apply to walking motions that are not well modeled as a wave-front error.

3 THE MACHINE LEARNING APPROACH

The core concept of our machine learning approach for detection is to use a sliding window search for the object feature, using image features extracted from the window and applying a classifier to those features. For different difficult environments we can learn different classifiers. We first review the learning and image features used.

3.1 Learning Techniques

Supervised learning is a machine learning approach that aims to estimate a classification function f from a *training data set*. Such a training data set consists of pairs of input values X and its desired outcomes Y [18]. Observed values in X are denoted by x_i, i.e., x_i is the i^{th} observation in X. Often, x is as simple as a sequence of numbers that represent some observed features. The number of variables or features

[1] `http://www.aoptix.com/`

in each $x \in X$ is p. Therefore, X is formed by N input examples (vectors) and each input example is composed by p features or variables.

The commonest output of the function f is a label (class indicator) of the input object under analysis. The learning task is to predict the function outcome of any valid input object after having seen a sufficient number of training examples.

In the literature, there are many different approaches for supervised learning such as Linear Discriminant Analysis, Support Vector Machines (SVMs), Classification Trees, and Neural Networks. We focus on an SVM-based solution.

3.2 PCA Features

Principle Components Analysis [19], that battle-worn method of statistics, is well suited to the dimensionality reduction of image data. Mathematically defined, PCA is an orthogonal linear transformation, which after transforming data leaves the greatest variance by any projection of data on the first coordinate (the *principal component*), and each subsequent level of variance on the following coordinates. For a data matrix X^T, after mean subtraction, the PCA transformation is given as

$$Y^T = X^T W = V\Sigma \tag{1}$$

where $V\Sigma W^T$ is the singular value decomposition of X^T. In essence, for feature detection, PCA provides a series of coefficients that become a feature vector for machine learning. Varying numbers of coefficients can be retained, depending on the energy level that provides the best detection resolution.

3.3 PCA + Learning Algorithm

A learning based feature detection approach allows us to learn over features gathered in the appropriate scenarios in which a recognition system will operate, including illumination, distance, pose, and weather (Fig. 1). By projecting a set of candidate pixels against a pre-computed PCA subspace for a particular condition, and classifying the resulting coefficients using a machine learning system yields an extremely powerful detection approach. The basic algorithm, depicted in Figure 2, begins with the results of the Viola-Jones face detector [20], implemented to return a face region that is symmetrical. With the assumption of symmetry, the face can be separated into feature regions, which will be scanned by a sliding window of a pre-defined size $w \times h$. Each positive marginal distance returned by an SVM classifier is compared against a saved maximum, with new maximums and corresponding x and y coordinates being saved. When all valid window positions are exhausted, the maximum marginal value indicates the candidate feature coordinate with the highest confidence. While for this work we are only interested in the eyes, we do note that

Fig. 1 The approach: build classifiers for different conditions, such as distance and illumination. While general enough for any feature that can be represented by a window of pixels, the eye is shown here, and in subsequent figures, as an example

the generic nature of the proposed approach allows for the detection of any defined feature.

The speed of such a sliding window approach is of concern. If we assume the window slides one pixel at a time, a 50×45 window (a window size suitable for eye detection on faces 160 pixels across) in an 80×60 potential feature region, 496 locations must be scanned. One approach to enhancing speed is through the use of multiple resolutions of feature regions. Figure 3 depicts this, with the full feature region scaled down by 1/4 as the lowest resolution region considered by the detector. The best positive window (if any) then determines a point to center around for the second (1/2 resolution) scale's search, with a more limited bounding box defined around this point for the search. The process then repeats again for the highest resolution. Presuming a strong classifier, the positive windows will cluster tightly around the correct eye region. A further enhancement to the algorithm is to determine the best positive window for the first row of the feature region where positive detections have occurred. From the x coordinate of this best window x_{best}, the scanning procedure can reduce the search space to $(x_{best} + c) - (x_{best} - c) + 1$ windows per row of the feature region, where c is some pixel constant set to ensure flexibility for the search region. c pixels will be searched on both the left and right sides of x_{best}. This approach does come with a drawback - the space requirement for

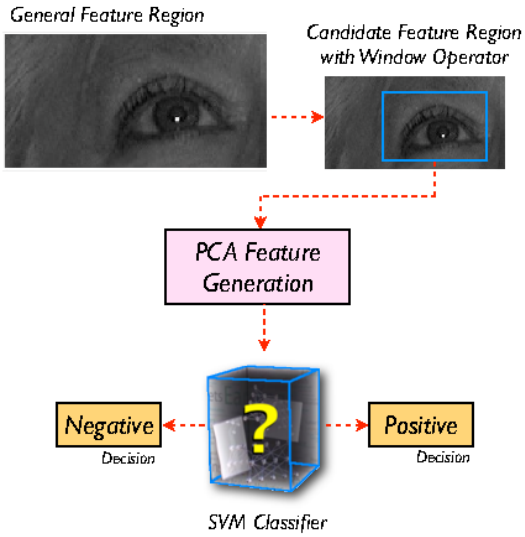


Fig. 2 The basic algorithm is straightforward. First, a feature region is isolated (using pre-set coordinates) from the face region returned by the face detector (separate from the feature detection). Next, using a pre-defined sliding window over the feature region, candidate pixels are collected. PCA feature generation is then performed using the pixels in the window. Finally, the PCA coefficients are treated as feature vectors for an SVM learning system, which produces the positive or negative detection result

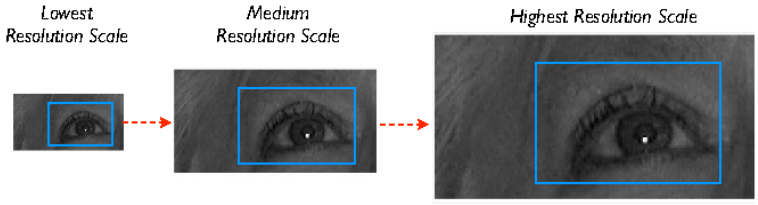


Fig. 3 The speed of sliding window approaches is always a concern. To increase computational performance, a multi-resolution approach can be used to reduce the area that must be scanned. While reducing time, this does increase the space requirement for PCA subspaces and SVM classifiers (number of features $\times$ number of scales)

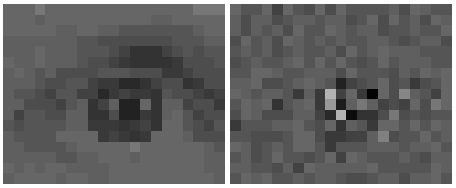
PCA subspaces and SVM classifiers increases by number of features $\times$ number of scales.

4 THE CORRELATION FILTER APPROACH

Correlation filters as considered in this work consist of Minimum Average Correlation Energy (MACE) filters [10], Unconstrained Minimum Average Correlation Energy (UMACE) filters [11], and our own Adaptive Average Correlation Energy (AACE) filters. These approaches produce a single correlation filter for a set of training images. For feature detection, these techniques produce a sharp correlation peak after filtering in the positive case, from which the correct coordinates for the feature can be derived (an example of this is shown in Figure 6). The variations among MACE, UMACE, and AACE are described below.

4.1 MACE Filter for Feature Detection

Synthesis of the Minimum Average Correlation Energy (MACE) filter began with cropping out 40×32 regions from our training data with the eye centered at coordinates (21,19). Figure 4(a) shows an example cropped eye from one of our training images.



(a) (b)

Fig. 4 Example cropped eye for MACE filter training (a) Impulse response from MACE filter (b)

The MACE filter specifies a single correlation value per input image, which is the value that should be returned when the filter is centered upon the training image. Unfortunately when more than 4-6 training images are used this leads to over fitting of the training data and decreases accuracy in eye detection. After cropping the eye region, it is transformed to the frequency domain using a 2D Fourier transform. Next the average of the power spectrum of all of the training images is obtained. Then MACE filter is synthesized using the following formula:

$$h = D^{-1}X(X'D^{-1}X)^{-1}u \tag{2}$$

where D is the average power spectrum of the N training images, X is a matrix containing the 2D Fourier transform of the N training images, and u is the desired filter output. Separate MACE filters were designed for both the left and right eyes.

The impulse response of the MACE filter for experiments shown in Figures 15 & 16 is shown in figure 4(b).

To incorporate a motion blur estimate into the MACE filter, an additional convolution operation must be executed prior to eye detection which can be performed on at run time on a per image basis. Finally, after the normalized cross correlation operation is performed the global maximum or peak location is chosen as the detected eye location in the original image with the appropriate offsets.

4.2 UMACE and AACE Filters for Feature Detection

Synthesis of our Adaptive Average Correlation Energy (AACE) filter is based on a UMACE filter. We start the filter design by cropping out regions of size 64×64 for the training data, with the eye centered at coordinates (32,32). After the eyes are cropped, each cropped eye region is transformed to the frequency domain using a 2D Fourier transform. Next, the average training images and the average of the power spectrum is calculated. The base UMACE filter for our AACE filter is synthesized using the following formula:

$$h = D^{-1}m \tag{3}$$

where D is the average power spectrum of the N training images, and m is the 2D Fourier transform of the average training image.

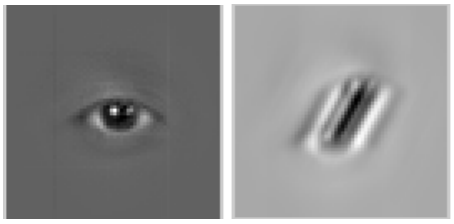
Separate filters were designed for both the left and right eyes. The UMACE filter is stored in its frequency domain representation to eliminate another 2D Fourier transform before the correlation operation is performed. Since we are performing the correlation operation in the frequency domain the UMACE filter had to be preprocessed by a Hamming window to help reduce the edge effects and impact of high frequency noise that is prevalent in the spectrum of low-light EMCCD imagery. Our experiments showed that windowing both the filter and input image decreased the accuracy of the UMACE eye detector. Since the UMACE filter is trained off line it was chosen as the input that was preprocessed by the Hamming window. One advantage of the UMACE filter over the MACE filter is that over-fitting of the training data is avoided by averaging the training images. Furthermore, we found that training data taken under ideal lighting conditions performed well for difficult detection scenarios when combined with an effective lighting normalization algorithm (discussed in section 5.4.1). This allows us to build an extremely robust filter that can operate in a wide array of lighting conditions, instead of requiring different training data for different lighting levels, as was the case with the machine learning based detector.

Furthermore, our motion blur estimate or point spread function (PSF) can be convolved into UMACE filter using only a point wise multiply of the motion blur Optical Transfer Function (OTF) and the UMACE filter. The resulting filter is what we call our Adaptive Average Correlation Energy (AACE) filter. The concept of the AACE filter is to take the UMACE filter, trained on good data, and adapt it, per

image, for the environmental degradations using estimates of blur and noise. The AACE filter is synthesized using the following formula:

$$h = (D^{-1}m) \otimes BlurOTF \tag{4}$$

Figure 5 shows the impulse response of a unblurred and motion blurrred AACE filter.



(a) (b)

Fig. 5 Impulse response of AACE filter: Unblurred (a); Motion Blurred (b).

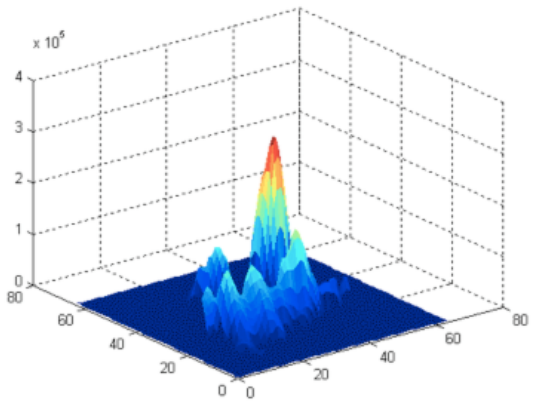


Fig. 6 Example Correlation Output with the Detected Eye Centered at Coordinates (40,36)

Finally, after the correlation operation is performed the global maximum or peak location is chosen as the detected eye location in the original image with the appropriate offsets. Figure 6 shows an example correlation output with the detected eye centered at coordinates (40,36).

5 EXPERIMENTS

Generating statistically significant datasets for difficult acquisition circumstances is a laborious and time consuming process. Capturing real world variables such as atmospheric turbulence, specific lighting conditions, and othe real world scenarios exacerbate the problem further. A specialized experimental setup called "photo-head" introduced by [5] showed using quality guided-synthetic data was a feasible evaluation technique for face recognition algorithm development.

5.1 The Photo-head Testing Protocol

In the setup described in that work, two cameras were placed 94ft and 182ft from a weather-proof LCD panel in an outdoor setting. The FERET data set was displayed on the panel at various points throughout the day, where it was re-imaged by the cameras over the course of several years. This unique re-imaging model is well suited to biometric experiments, as we can control for distance, lighting and pose, as well as capture statistically meaningful samples in a timely fashion. Further, it allows for reproducible experiments by use of standard data sets that are re-imaged.

In our own setup, instead of imaging an LCD panel, we used a Mitsubishi PK10 LCD pocket projector, which has a resolution of 800×600 pixels and outputs 25 ANSI Lumens, to project images onto a blank screen. The experimental apparatus was contained in a sealed room, where lighting could be directly controlled via the application of polarization filters to the projector. The camera used for acquisition was a SI-1M30-EM low-light EMCCD unit from FLIR Systems. At its core, this camera utilizes the TI TX285SPD-B0 EMCCD sensor, with a declared resolution of $1,004 \times 1,002$ (the effective resolution is actually $1,008 \times 1,010$). To simulate distance, all collected faces were roughly 160 pixels in width (from our own work in long distance acquisition, this is typical of what we would find at 100M with the appropriate optics). Photo-head images can be seen in Figure 9.

In order to assess and adjust the light levels of the photo-head imagery, we directly measure the light leaving the projected face in the direction of the sensor - its *luminance*. The candela per square meter ($\frac{cd}{m^2}$) is the SI unit of luminance; nit is a common non-SI name also used for this unit (and used throughout the rest of this paper). Luminance is valuable because it describes the "brightness" of the face and does not vary with distance. For our experiments, luminance is the better measure to assess how well a particular target can be viewed - what is most important for biometric acquisition. More details on this issue of light and face acquisition can be found in [21].

5.2 Evaluation of Machine Learning Approach

In order to assess the viability of the detector described in Section 3.3, a series experiments under very difficult conditions was devised. First, using the photo-head methodology of Section 5.1, a subset of the CMU PIE [22] data set was re-imaged in a controlled (face sizes at approximately the same width as what we would collect at 100M), dark indoor setting (0.043 - 0.017 face nits). Defined feature points are the eyes, with a window size of 45×35 pixels. For SVM training, the base positive set consisted of 250 images $\times$ (8 1-pixel offsets from the ground-truth + ground-truth point), for each feature. The base negative set consisted of 250 images $\times$ 9 pre-defined negative regions around the ground-truth, for each feature. The testing set consisted of 150 images per feature. The actual data used to train the PCA subspaces and SVM classifiers varies by feature, and was determined experimentally based on performance. For the left eye, 1,000 training samples were provided for subspace training, and for the right eye, 1,200 samples were provided. The experiments presented in this section are tailored to assess accuracy of the base technique, and are performed at the highest resolution possible, with the window sliding 1 pixel at a time.

The results for eye detection are shown in Figures 7 and 8. On each plot, the x axis represents the pixel tolerance as a function of distance from the ground-truth for detection, and the y axis represents the detection percentage at each tolerance. The proposed detection approach shows excellent performance for the photo-head imagery. For comparison, the performance of a leading commercial detector (chosen for its inclusion in a face recognition suite that scored at or near the top of every test in FRVT 2006), is also plotted. The proposed detection approach clearly outperforms it till both approaches start to converge after the pixel tolerance of 10. Examples of the detected feature points from the eye comparison experiment are shown in Figure 9.

Even more extreme conditions were of interest for this research. Another photo-head set was collected based on the FERET [23] dataset between 0.0108 - 0.002 nits. For an eye feature (left is shown here), a window of 50×45 was defined. The Gallery subset was used for training, with a subspace of 1100 training samples, and a classifier composed of 4200 training samples (with additional images generated using the perturbation protocol above). For testing, all of the FAFC subset was submitted to the detector. The results of this experiment are shown in Figure 10; the commercial detector is once again used for comparison. From the plot, we can see the commercial detector failing nearly outright at these very difficult conditions, while the proposed detector performs rather well.

Blur is another difficult scenario that we have looked at. For this set of experiments, we produced a subset of images from the FERET data set (including the ba, bj, and bk subsets) for three different uniform linear motion models of blur: blur length of 15 pixels, at an angle of 122 degrees; blur length of 17 pixels, at an angle of 59 degrees; blur length of 20 pixels, at an angle of 52 degrees. Sample images from each of these sets are shown in Figure 11. The classifier for detection was trained using 2000 base images of the left eye (split evenly between positive and

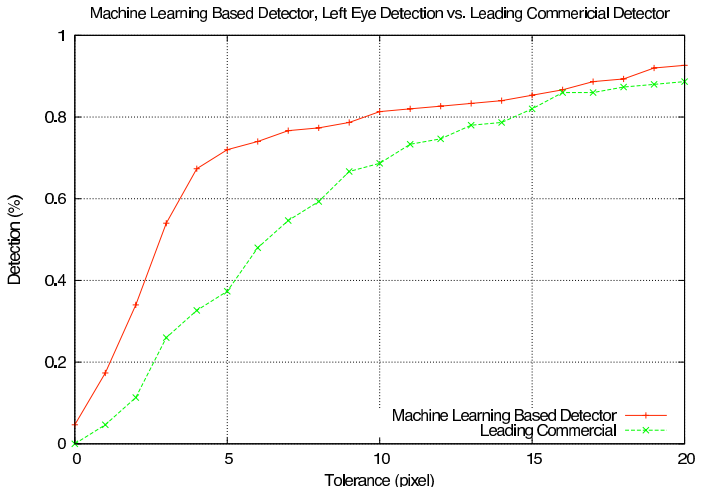


Fig. 7 Performance of the proposed machine learning based detector against a leading commercial detector for the left eye under dark conditions. The machine learning based detector clearly outperforms the commercial detector

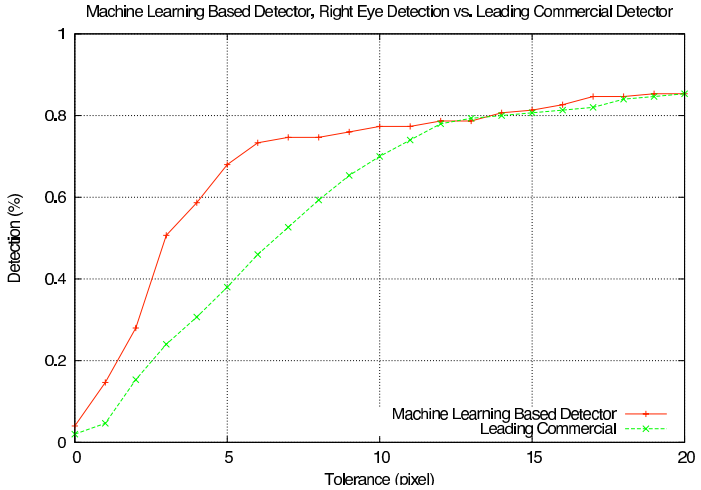


Fig. 8 Performance of the proposed machine learning based detector against a leading commercial detector for the right eye under dark conditions. Results are similar to the left eye in figure 7.

negative training samples), derived from 112 base images (again, additional images were generated using the perturbation protocol above) at the blur length of 20 pixels, at an angle of 52 degrees. The subspace was trained on 1000 positive images, with the same blur model. The testing set consisted of 150 images, with the left eye as the feature target for each of the three blur models. The results for this experiment are

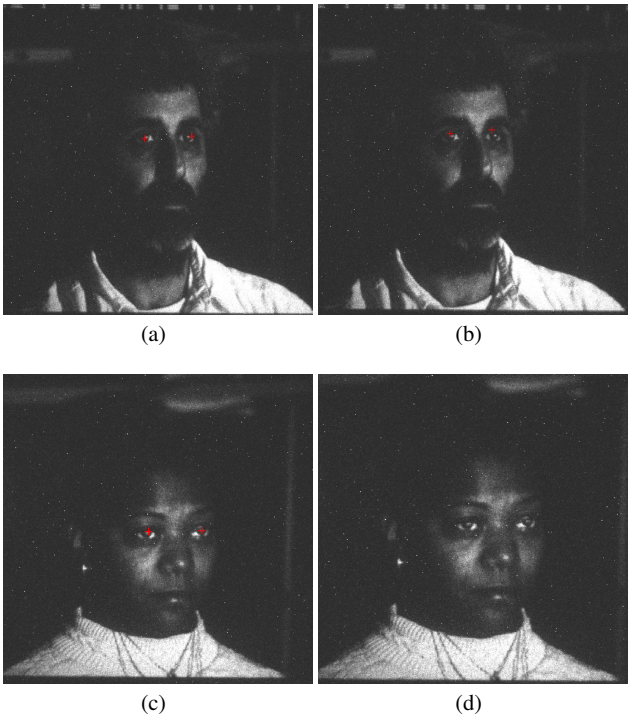
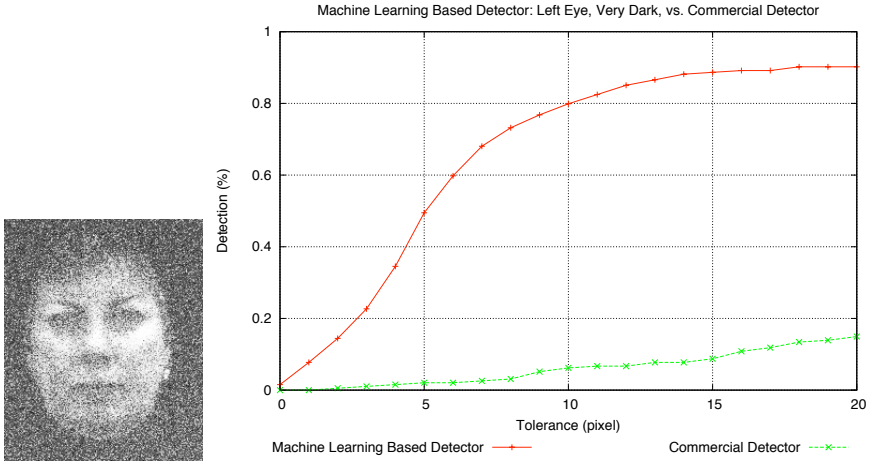


Fig. 9 Qualitative results for the proposed machine learning based detector (left) and a leading commercial detector (right) for comparison. The commercial detector is not able to find any eyes in image d

shown in Figure 12. From this figure, we can see that the machine learning based approach has a slight advantage over the commercial detector for the blur length of 20 pixels - the blur model it was trained with. For testing with the other blur models, performance is acceptable, but drops noticeably. Thus, we conclude that incorrect blur estimations can negatively impact this detection approach.

Reduced resolution imagery (face sizes $\leq 90 \times 90$ pixels), is another difficult scenario that we have explored. The performance of most face recognition algorithms degrades substantially whenever the input images are of low resolution or size, which is often the case whenever the images are taken by a surveillance camera in an uncontrolled setting, since these algorithms were designed and developed with high or average quality images at close ranges ≤ 3 meters. Recent work from the face recognition community is addressing the issue of recognizing subjects from low quality or reduced resolution images [24, 25, 26]. However, accurate eye detection is still vital to provide optimal performance when using these reduced resolution face recognition algorithms. This set of experiments was designed to examine how



(a) (b)

Fig. 10 Results comparing the machine learning based detector to a leading commercial detector, for the left eye, with very dark imagery (0.0108 - 0.002 nits). A sample image (a) is provided to signify the difficulty of this test (histogram equalized here to show "detail"). The commercial detector fails regularly under these conditions

(a) (b) (c)

Fig. 11 Examples of blurred imagery for three different blur models used for experimentation. (a) Blur length of 15 pixels, at an angle of 122 degrees (b) Blur length of 17 pixels, at an angle of 59 degrees (c) Blur length of 20 pixels, at an angle of 52 degrees

our machine learning based detector performs on the same data set at full resolution and at a reduced resolution; down sampled by $2\times$ in each direction.

For this set of experiments, we again used a subset of images from the FERET data set (including the ba, bj, and bk subsets) at the full and reduced resolution, where images were down sampled by 2 $\times$ in each direction. Sample images from each of these sets are shown in Figure 13. The classifier for eye detection was trained using 2000 base images of the left eye (split evenly between positive and negative training samples), derived from 200 base images (additional images were generated using the perturbation protocol). The subspace was trained with the 1000 positive

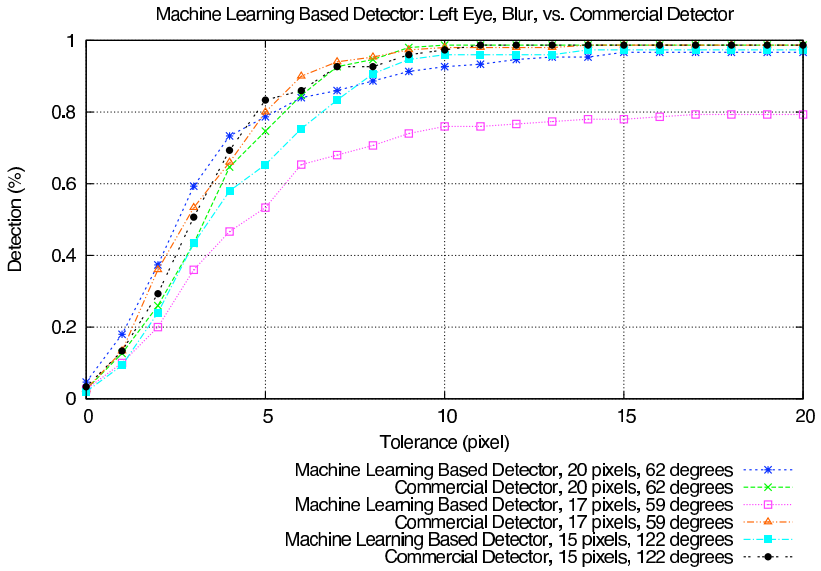
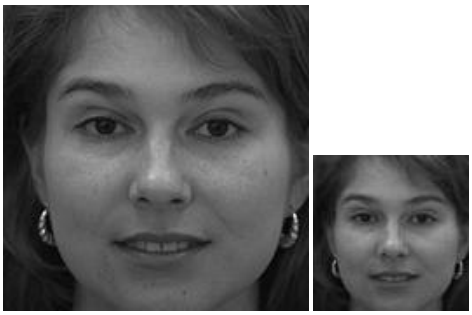


Fig. 12 Results comparing the machine learning based detector to a leading commercial detector, for the left eye, with three varying degrees of blur. The detector was trained using the blur length of 20 pixels at an angle of 52 degrees



(a) (b)

Fig. 13 Examples of full and reduced resolution imagery used for experimentation. Sample image full resolution 176×176 (a). Sample image reduced resolution 89×89 (b)

images. The testing set consisted of 150 images, with the left eye as the feature target for each of the models. The results for this experiment are shown in Figure 14. From this figure, we can see that the machine learning based approach has a slight advantage over the commercial detector for pixel tolerances < 5 (following this both detectors converge).

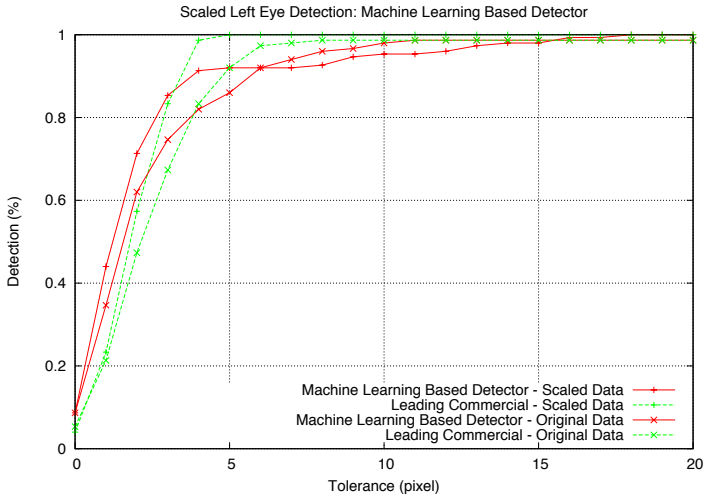


Fig. 14 Results comparing the machine learning based detector to a leading commercial detector, for the left eye, for full and reduced resolution images. The machine learning based detector outperforms the commercial detector for pixel tolerances < 5 (following this both detectors converge)

5.3 Evaluation of Correlation Approach

The experiments performed for the correlation approach are identical to the ones we performed for the machine learning approach, with the following training details. The MACE filters used in Figures 15 & 16 were trained with 6 eye images, while the MACE filter for Figure 17 used 4 training images (these values were determined experimentally, and yield the best performance). For the experiments of Figure 15 & 16, the AACE filter was synthesized with 266 images, for the experiment of Figure 17, the filter was synthesized with 588 images. For the AACE filter used in the experiment of Figure 18, the filter was synthesized with 1500 images, incorporating the exact same blur model as the machine learning experiments into the convolution operator. Furthermore, the training data for the experiments in Figures 15, 16 & 17 used images taken under ideal lighting conditions. For the AACE filter used in the experiment of Figure 19, the filter was synthesized with the same 1500 images for the motion blur experiment though no PSF model was incorporated into the AACE filter.

Comparing the AACE approach to the machine learning approach, the correlation filter detector shows a significant performance gain over the learning based detector on blurry imagery (Figure 12 vs. Figure 18). What can also be seen from our experiments is a stronger tolerance for incorrect blur estimation, with the blur length of 17 pixels, 59 degrees performing just as well as the training blur model; this was not the case with the machine learning based detector. In all other experiments, the

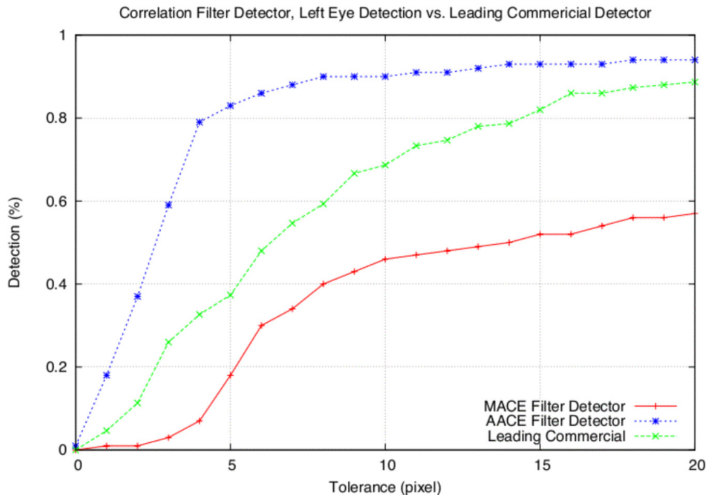


Fig. 15 Performance of the correlation filter detectors against a leading commercial detector for the left eye under dark conditions

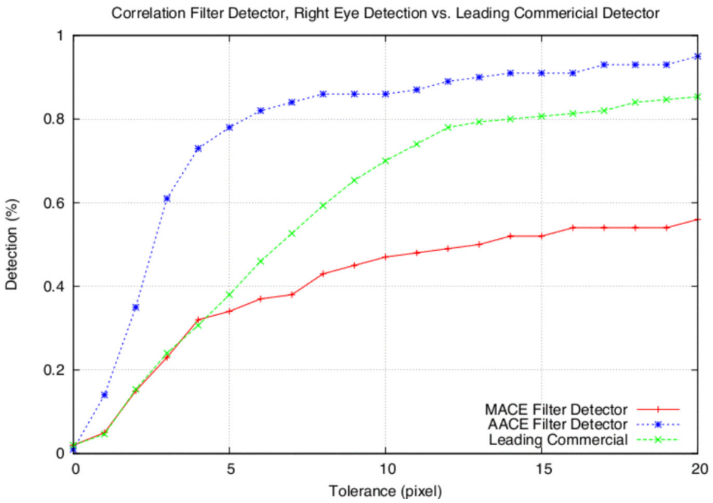


Fig. 16 Performance of correlation filter detectors against a leading commercial detector for the right eye under dark conditions

AACE filter detector produced a modest performance gain over the machine learning based detector. The performance of MACE was poor for all test instances that it was applied to.

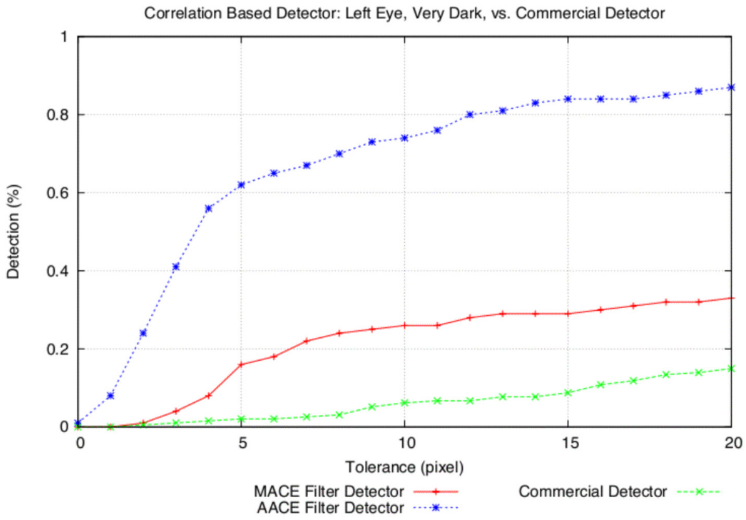


Fig. 17 Results for the correlation filter detectors for the left eye, with very dark imagery (0.0108 - 0.002 nits)

5.4 Methods to Improve Correlation Approach

5.4.1 Lighting Normalization

In addition to using multiple AACE eye filter models for different lighting situations, we decided to implement and test a lighting normalization algorithm to see whether it would increase the accuracy of the eye detector. A key motive for using lighting normalization in conjunction with our correlation eye detector came from some of our daytime experiments where the faces had extreme shadows and gradients on them. These shadows and gradients on the face were causing the eye detector to improperly identify the position of the eye as shown below in Figure 20.

Our lighting normalization algorithm, is presented below. We are currently using a modified version of the Self-Quotient illumination (SQI) lighting normalization algorithm. Self-Quotient illumination (SQI) normalization is based on the work of [4]. The SQI image is formed by dividing the original face image $f(x,y)$ with the original image convolved with a Gaussian function that acts as a smoothing kernel function $S(x,y)$.

$$Q(x,y) = \frac{f(x,y)}{S(x,y)} = \frac{f(x,y)}{G(x,y) \otimes f(x,y)} \tag{5}$$

The subsequent task of the lighting normalization method is to normalize $Q(x,y)$ to have pixel intensity between 0 and 1, and to increase the contrast of the image by applying linear transformation function.

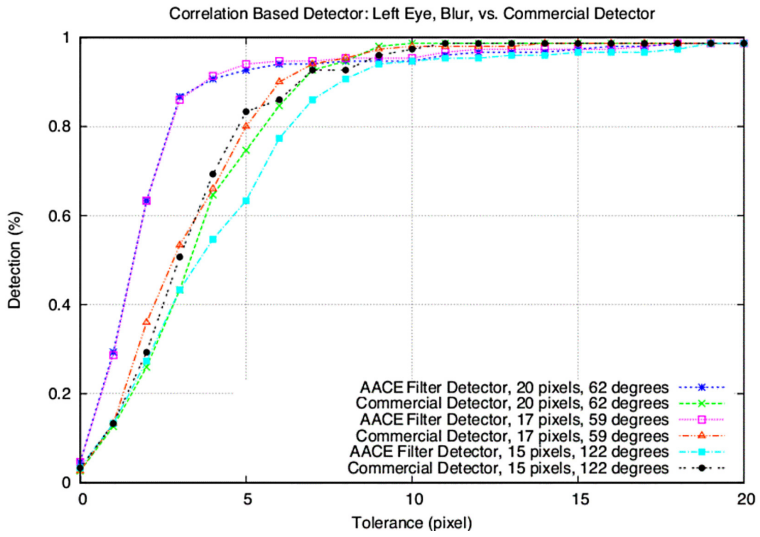


Fig. 18 Results comparing the AACE correlation filter detector to a leading commercial detector, for the left eye, with three varying degrees of blur. The filters were trained using the blur length of 20 pixels, at an angle of 52 degrees

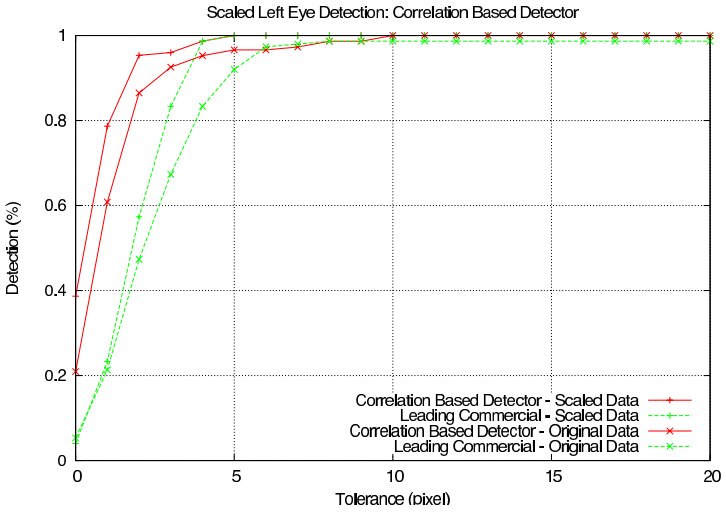
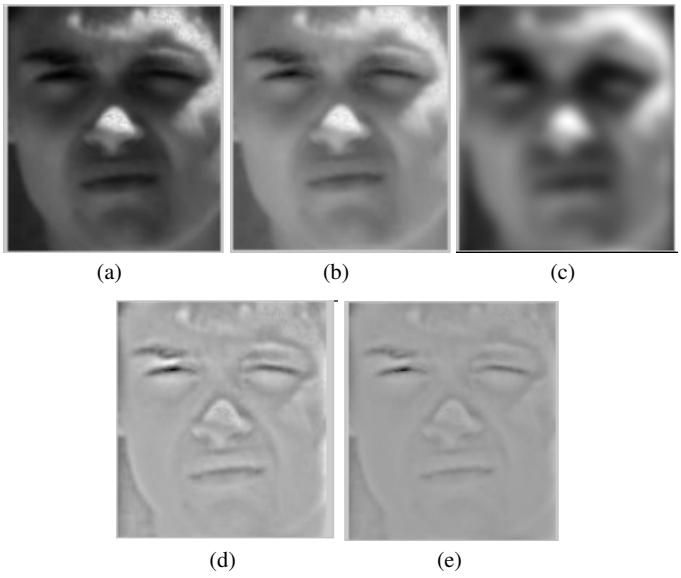


Fig. 19 Results comparing the AACE correlation filter detector to a leading commercial detector, for the left eye, for full and reduced resolution images

(a) (b)

Fig. 20 Output of Eye Detector Without Lighting Normalization; the right eye position is not properly identified (a). Output of eye detector with lighting normalization; the right eye position is properly identified (b)



(a) (b) (c)

(d) (e)

Fig. 21 Lighting Normalization Algorithm with Example Daytime Image (a) Original Image (b) Gamma Correction of Image (c) Smoothed Image (d) Quotient Image (e) Normalized Quotient Image

$$Q'(x,y) = \frac{Q(x,y) - Q_{min}}{Q_{max} - Q_{min}} \tag{6}$$

$$Q_{norm}(x,y) = 1 - e^{-\frac{Q'(x,y)}{E(Q'(x,y))}} \tag{7}$$

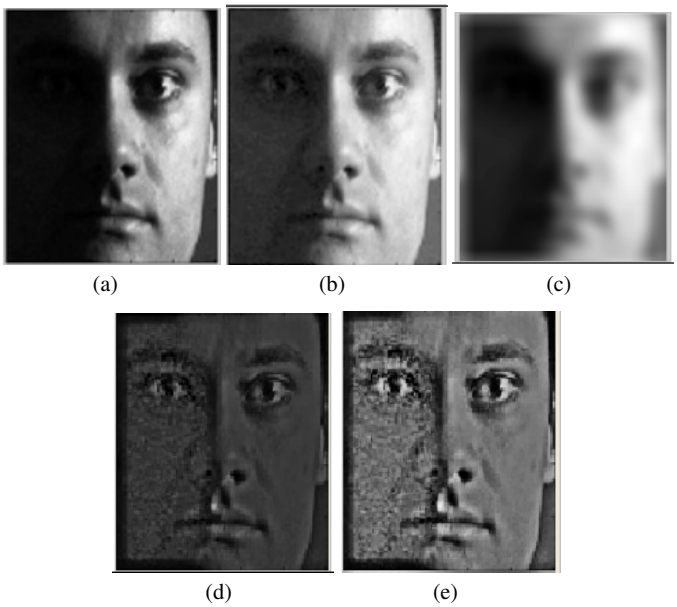


Fig. 22 Lighting Normalization Algorithm with Example Low-Light Image (a) Original Image (b) Gamma Correction of Image (c) Smoothed Image (d) Quotient Image (e) Normalized Quotient Image

where Q_{max} and Q_{min} are maximum and minimum values of Q respectively, and $E(.)$ is a mean value. Therefore, Q_{norm} is a normalized Gaussian quotient image and is used as an image for eye detection as shown below in Figures 21 and 22.

5.4.2 Eye Location Perturbations

A known problem with correlation based eye detectors is that they will also show a high response to eyebrows, nostrils, dark rimmed glasses, and strong lighting such as glare from eye glasses and return these points as the coordinates of the eye. Through our analysis of the problem we have discovered that when an invalid location has the highest correlation peak value, a second or third correlation peak with a value slightly less than the highest peak is usually the true location of the eye. Therefore, our eye detection algorithm has been modified to search for multiple correlation peaks on each side of the face and then determine which correlation peak is the true location of the eye. Once the initial correlation output is returned it is thresholded at 80% of the maximum value to eliminate all but the salient structures in the correlation output. A unique label is then assigned to each structure using connected component labeling [27]. The location of the maximum peak within each label is

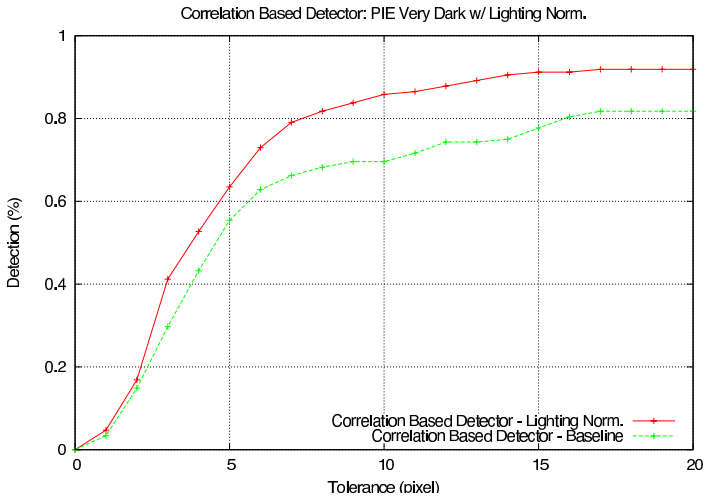


Fig. 23 Results for the correlation filter detector for the left eye, with very dark imagery (0.043 - 0.017 nits) with and without lighting normalization

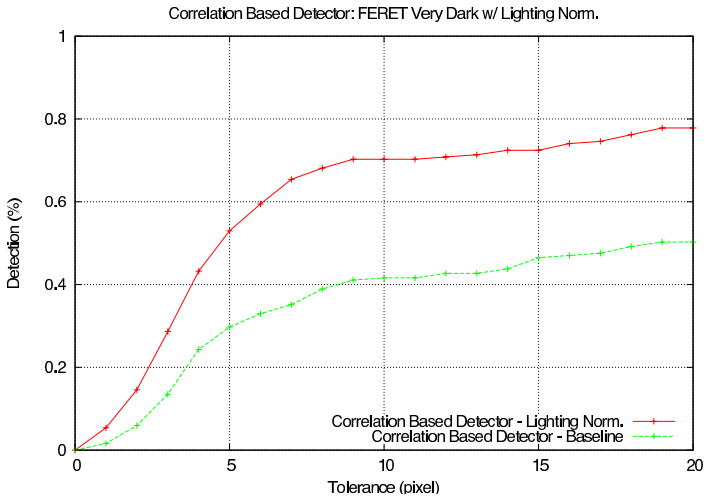


Fig. 24 Results for the correlation filter detector for the left eye, with very dark imagery (0.0108 - 0.002 nits) with and without lighting normalization

then located and returned as a possible eye location. This process is repeated for both sides of the face.

Our ultimate goal is to determine the location of the left and right eye and then send the input image and the eye locations to a geometric normalization algorithm. However, we are taking a different approach by sending all of the initial eye locations to the geometric normalization algorithm and then determining the “best”

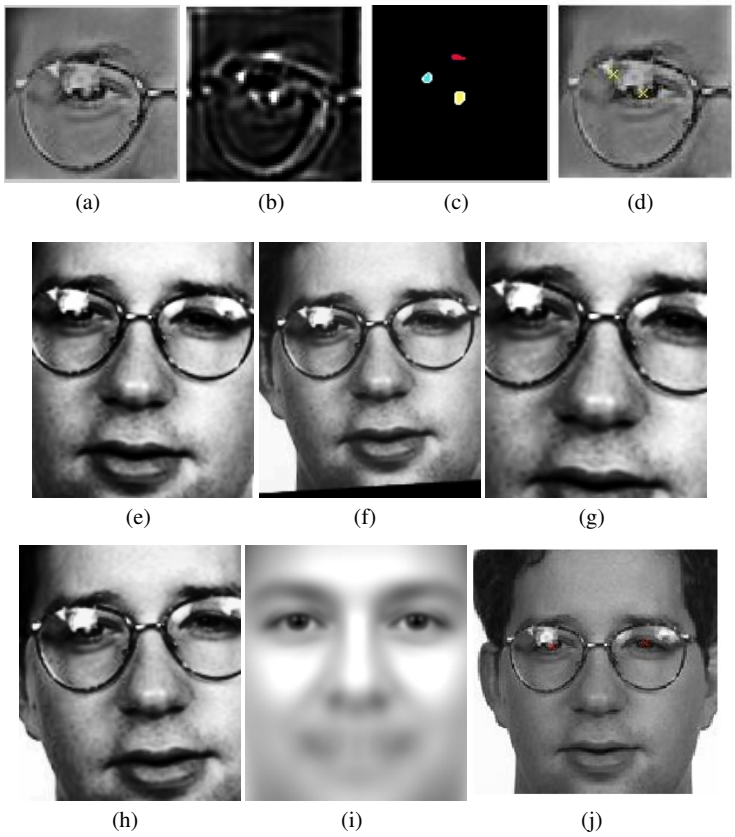


Fig. 25 (a) Cropped left eye area (b) Correlation output (c) Connected components image derived from thresholded correlation output (d) Cropped left eye area with top two initial eye locations returned (e-h) Image perturbations using top two initial left and right eye locations (i) "Average Face" (j) Final eye coordinates returned based on top score using perturbation algorithm

geometrically normalized image from all of the normalized images. Geometric normalization is a vital step in our face recognition pipeline since it reduces the variation between gallery and probe images. The geometrically normalized image is of uniform size and if the input eye coordinates are correct the output image will contain a face chip with uniform orientation. All of the geometrically normalized images are compared against an "average" face using normalized cross-correlation. Our "average" face was formed by first geometrically normalizing and then averaging all of the faces from the FERET data set [23]. Normalized cross-correlation is only performed on a small region around the center of the image. The left and right (x,y) eye coordinates corresponding to the image with the highest similarity are returned as the true eye coordinates. Additionally, since we have already performed geometrically normalization this step does not need to be performed again in our

pipeline. A summary of the new algorithm is shown in Figure 25. Only the top two eye coordinates were considered on each side of the face.

5.4.3 Evaluation of Eye Location Perturbations

(a) (b)

Fig. 26 Example imagery from SCface - Surveillance Cameras Face Database

To evaluate the performance of the correlation approach using the eye location perturbation algorithm presented in Figure 5.4.2 we performed an experiment using 128 images from the SCface - Surveillance Cameras Face Database [25]. The images in the SCface database are taken from various surveillance cameras with uncontrolled lighting and the images are of various quality and resolution. Example images from our test set are shown in Figure 26. For the correlation filter used in the experiment of Figure 27, the filter was synthesized with the same 1500 images from the motion blur experiment with no PSF model being incorporated into the AACE filter. The lighting normalization algorithm presented in 5.4.1 was used on the images prior to eye detection. Only two (x,y) eye coordinates were considered on each side of the face for this experiment. The results for eye detection are shown in Figure 27. The proposed detection approach shows a moderate performance gain for the difficult imagery.

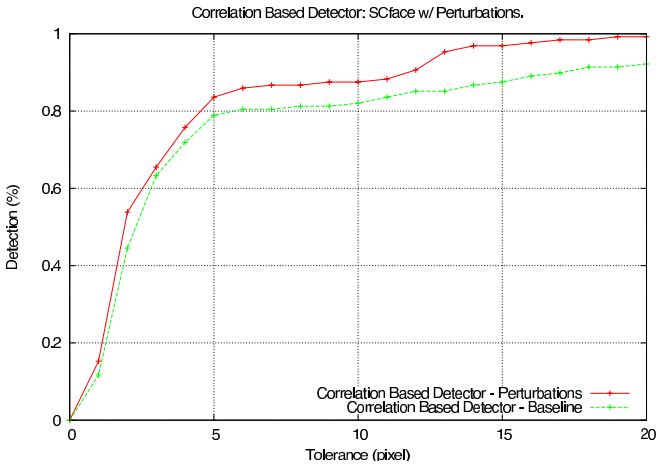


Fig. 27 Results for the correlation filter detectors with and without using eye location perturbation algorithm

6 CONCLUSIONS

As face recognition moves forward, difficult imagery becomes a primary concern. But before we can even attempt face recognition, we often need to perform some necessary pre-processing steps, including geometric normalization and facial feature localization, with the eyes providing the necessary reference points. Thus, in this paper, we have concentrated on the eye detection problem for unconstrained environments. First, we introduced an EMCCD approach for low-light acquisition, and subsequently described an experimental protocol for simulating low-light conditions, distance, pose variation and motion blur. Next, we described two different detection algorithms: a novel machine learning based algorithm and a novel adaptive correlation filter based algorithm. Finally, using the data generated by our testing protocol, we performed a thorough series of experiments incorporating the aforementioned conditions. Both approaches show significant performance improvement over a leading commercial eye detector.

Comparing both approaches, our new AACE correlation filter detector shows a significant performance gain over the learning based detector on blurry imagery, and a moderate performance gain on low-light imagery. Our lighting normalization results showed that we could build a AACE correlation filter that can operate in a wide array of lighting conditions, instead of requiring different training data for different lighting levels. The perturbation algorithm showed that we could use multiple eye estimates to ultimately help select the real eye locations. Based on the presented results, we conclude that both approaches are suitable for the problem at hand - the choice of one as a solution can be based upon implementation requirements. As far as we know, this is the first study of feature detection under a multitude of difficult

acquisition circumstances, and its results give us confidence for tackling the next steps for unconstrained face recognition.

Acknowledgements This work was supported by ONR STTR Biometrics in the Maritime Domain, (Award Number N00014-07-M-0421), ONR MURI (Award Number N00014-08-1-0638), São Paulo Research Foundation, FAPESP (Award Number 2010/05647-4) and Unicamp's PAPDIC program (Award Number 34/010). Portions of the research in this paper use the SCface database of facial images. Credit is hereby given to the University of Zagreb, Faculty of Electrical Engineering and Computing for providing the database of facial images.

References

1. P.J. Phillips and Y. Vardi. Efficient illumination normalization of facial images. *Elsevier Pattern Recognition Letters (PRL)*, 17(8):921–927, July 1996.
2. Terrence Chen, Wotao Yin, Xiang Sean Zhou, Dorin Comaniciu, and Thomas S. Huang. Total variation models for variable lighting face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(9):1519–1524, 2006.
3. Yi Yao, Besma R. Abidi, Nathan D. Kalka, Natalia A. Schmid, and Mongi A. Abidi. Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Elsevier Computer Vison and Image Understanding (CVIU)*, 111(2):111–125, 2008.
4. B. Heflin, B. Parks, W. Scheirer, and T. Boult. Single image deblurring for a real-time face recognition system. In *IEEE Industrial Electronics Society (IECON)*, 2010.
5. T. Boult, W. Scheirer, and R. Woodworth. Faad: Face at a distance. In *SPIE Defense and Security Symposium*, April 2008.
6. B. Leite, E. Pereira, H. Gomes, L. Veloso, Santos C., and Carvalho J. A learning-based eye detector coupled with eye candidate filtering and pca features. In *Intl. Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2007.
7. C. Shuo and C. Liu. Aeye detection using color information and a new efficient svm, 2010.
8. Peng Wang, Matthew B. Green, Qiang Ji, and James Wayman. Abstract automatic eye detection and its validation. In *IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
9. Lizuo Jin, Xiaohui Yuan, Shinichi Satoh, Jiuxian Li, and Liangzheng Xia. A hybrid classifier for precise and robust eye detection. In *Intl. Conference on Pattern Recognition (ICPR)*, pages 731–735, Washington, DC, USA, 2006. IEEE Computer Society.
10. Abhijit Mahalanobis, B. V. K. Vijaya Kumar, and David Casasent. Minimum average correlation energy filters. *Appl. Opt.*, 26(17):3633–3640, 1987.
11. Marios Savvides and B.V.K. Vijaya Kumar. Efficient design of advanced correlation filters for robust distortion-tolerant face recognition. In *IEEE Intl. Conference on Advanced Video and Signal Based Surveillance*, page 45, Los Alamitos, CA, USA, 2003. IEEE Computer Society.
12. Marios Savvides, B.V.K. Vijaya Kumar, and P.K. Khosla. "corefaces" - robust shift invariant pca based correlation filter for illumination tolerant face recognition. In *IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 834–841, Los Alamitos, CA, USA, 2004. IEEE Computer Society.
13. M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, and B.V.K. Vijayakumar. Partial and holistic face recognition on frgc-ii data using support vector machine. In *Computer Vision and Pattern Recognition Workshop*, page 48, Los Alamitos, CA, USA, 2006. IEEE Computer Society.
14. R. Brunelli and T. Poggio. Template matching: matched spatial filters and beyond. *Pattern Recognition*, 30:751–768, 1997.

15. D.S. Bolme, B.A. Draper, and J.R. Beveridge. Average of synthetic exact filters. In *IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2105–2112, Los Alamitos, CA, USA, 2009. IEEE Computer Society.
16. R. Tyson. *Introduction to adaptive optics*. SPIE The Intl. Society for Optical Engineering, 2000.
17. Joseph Carroll, Daniel C. Gray, Austin Roorda, and David R. Williams. Recent advances in retinal imaging with adaptive optics. *Opt. Photon. News*, 16(1):36–42, 2005.
18. C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1st edition, 2006.
19. L. Smith. A tutorial on principal components analysis, 2002.
20. P. Viola and M. Jones. Robust real-time face detection. *Intl. Journal of Computer Vision (IJCV)*, 57(2):137–154, 2004.
21. T. Boult and W. Scheirer. Long range facial image acquisition and quality. In *Biometrics for Surveillance and Security. Edited by Tistarelli, M.; Li S. ; Chellappa R.* Springer-Verlag, 2009.
22. T. Sim, S. Baker, and Bsat M. The cmu pose, illumination, and expression (pie) database. In *Intl. Conference on Automatic Face and Gesture Recognition (FG)*, 2002.
23. P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(10), 2000.
24. A. Sapkota, B.C. Parks, W.J. Scheirer, and T.E. Boult. Face-grab: Face recognition with general region assigned to binary operator. In *IEEE CVPR Workshop on Biometrics*, volume 1, pages 82–89, Los Alamitos, CA, USA, 2010. IEEE Computer Society.
25. Mislav Grgic, Kresimir Delac, and Sonja Grgic. SCface a surveillance cameras face database. *Multimedia Tools and Applications*, pages 1–17–17, October 2009.
26. V.N. Iyer, S.R. Kirkbride, B.C. Parks, W.J. Scheirer, and T.E. Boult. A taxonomy of face-models for system evaluation. In *IEEE Analysis and Modeling of Faces and Gestures (AMFG)*, pages 63–70, 2010.
27. L. G. Shapiro and G. C. Stockman. *Computer Vision*. Prentice Hall, Englewood-Cliffs NJ, 2001.